

Result No.	Query			Description		
	Score	Match	Length	ID	DB	Description
1	44	19.7	80246	4	US-09-078-294-4	Sequence 4, Appli
2	44	19.7	80595	4	US-09-078-294-3	Sequence 3, Appli
c 3	34.6	15.5	6038	4	US-09-305-639-4	Sequence 4, Appli
c 4	34.6	15.5	7622	4	US-09-305-639-1	Sequence 1, Appli
5	33.2	14.9	1160	1	US-08-006-082A-4	Sequence 4, Appli
6	33.2	14.9	1264	1	US-08-006-082A-1	Sequence 1, Appli
c 7	32.8	14.7	1951	4	US-09-058-947A-12	Sequence 12, Appli
c 8	32.8	14.7	1883	1	US-08-383-756-3	Sequence 3, Appli
c 9	32.8	14.7	1983	2	US-08-460-898-3	Sequence 3, Appli
c 10	32.8	14.7	3722	4	US-09-058-947A-1	Sequence 1, Appli
c 11	32.4	14.5	51952	3	US-08-947-823-1	Sequence 1, Appli
c 12	32	14.3	4254	2	US-08-443-639-7	Sequence 7, Appli
c 13	32	14.3	8920	2	US-08-446-855A-1	Sequence 1, Appli
c 14	32	14.3	8920	4	US-09-150-741-1	Sequence 1, Appli
15	31.4	14.1	480	1	US-08-282-581-4	Sequence 4, Appli
16	31.4	14.1	480	1	US-08-550-544-4	Sequence 4, Appli
c 17	31.4	14.1	1601	3	US-08-817-188-3	Sequence 3, Appli
c 18	31.4	14.1	2799	2	US-08-874-185-42	Sequence 42, Appli
19	31.4	14.1	3627	1	US-08-104-072B-6	Sequence 6, Appli
20	31.4	14.1	3627	1	US-08-351-413-7	Sequence 7, Appli
21	31.4	14.1	3627	2	US-09-025-583-7	Sequence 7, Appli
c 22	31.2	14.0	19124	2	US-08-487-826B-13	Sequence 13, Appli
c 23	30.8	13.8	5993	4	US-09-383-630-1	Sequence 1, Appli
c 24	30.8	13.8	5993	4	US-09-383-630-2	Sequence 2, Appli
c 25	30.6	13.7	480	1	US-08-282-581-4	Sequence 4, Appli
c 26	30.6	13.7	480	1	US-08-550-544-4	Sequence 4, Appli
27	30.6	13.7	1890	6	5312912-3	Patent No. 5312912





```
; FILING DATE: 16-JUN-94
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10814
; FILING DATE: 29-OCT-93
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 07/968,971
; FILING DATE: 30-OCT-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; US-08-383-756-3

Query Match 14.7%; Score 32.8; DB 1; Length 1983;
Best Local Similarity 54.0%; Pred. No. 0.95;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 49 cctgctgcttaccatggatttctcgcgtactattgtgtatagctgtatataaccgtataa 108
Db 1718 CGTGTGCTGGCGCTGCAGCTGCTCCCAATATATATATATATATATATATATATA 1659

Qy 109 tgcacatatatttatgaacatgattgaattgctgtgagttgttctcgcgtaaagatt 168
Db 1658 TATATATATATATATATATATATACATTTTGGCTATGCATTATTGCTGGAGGAATATCA 1599

Qy 169 tcaa 172
Db 1598 TCAA 1595

RESULT 9
US-08-460-898-3/C
; Sequence 3, Application US/08460898
; Patent No. 5850022
; GENERAL INFORMATION:
; APPLICANT: Dehesh, Katayoon
; APPLICANT: Voelker, Toni Alois
; APPLICANT: Hawkins, Deborah
; APPLICANT: Davies, Huw Maelor
; TITLE OF INVENTION: Production of Myristate in Plant Cells
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1(a)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,898
; FILING DATE: 05-JUN-1995
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/383,756
; FILING DATE: 02-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/361,695
; FILING DATE: 16-JUN-94
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10814
; FILING DATE: 29-OCT-93
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 07/968,971
; FILING DATE: 30-OCT-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 111-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; US-08-460-898-3

Query Match 14.7%; Score 32.8; DB 2; Length 1983;
Best Local Similarity 54.0%; Pred. No. 0.95;
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 49 cctgctgcttaccatggatttctcgcgtactattgtgtatagctgtatataaccgtataa 108
Db 1718 CGTGTGCTGGCGCTGCAGCTGCTCCCAATATATATATATATATATATATATATA 1659

Qy 109 tgcacatatatttatgaacatgattgaattgctgtgagttgttctcgcgtaaagatt 168
Db 1658 TATATATATATATATATATATATACATTTTGGCTATGCATTATTGCTGGAGGAATATCA 1599

Qy 169 tcaa 172
Db 1598 TCAA 1595

RESULT 10
US-09-058-947A-1/C
; Sequence 1, Application US/09058947A
; Patent No. 6274790
; GENERAL INFORMATION:
; APPLICANT: Kunst et al.
; TITLE OF INVENTION: Nucleic Acids Encoding Plant Enzyme
; TITLE OF INVENTION: Involved In Very Long Chain Fatty Acid Synthesis
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell
; ADDRESSEE: Leigh & Winston, LLP
; STREET: One World Trade Center, Suite
; STREET: 1600, 121 S.W. Salmon Street
; CITY: Portland
; STATE: OR
; COUNTRY: USA
; ZIP: 97204-2988
; COMPUTER READABLE FORM:
```



Query Match 14.5%; Score 32.4; DB 3; Length 51952;  
Best Local Similarity 47.5%; pred. No. 2.6;

Qy	5	cataattataattataattataataactgaaccatgggtccccctgctgtttacatg	64
Db	13606	CATTCTAACGTATCTATAAAAAATTCACATGAACACACACGTGTGAAGCGCATCTTCAGAAA	13665
Qy	65	gatttcgcgcactattgttatcacggtgatataataccgtataatcgtgatcatatatttata	124
Db	13666	AACCTAGTGTA	13725
Qy	125	tgaacatgattaaagctctggagtggtctctcatccgtaagaggtttcgaatgtaagtgtg	184
Db	13726	TATATATATATATATATATATATATATATATATCTTATTAATAAAGAAATGCTCTTATTTTCATTTTT	13785
Qy	185	aagagtcacaaaccccaaatcat	206
Db	13786	AATCTGGTTAAAAAAGAAATAT	13807

```

RESULT 12
US-08-443-639-7/c
; Sequence 7, Application US/08443639
; Patent No. 5981843
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: Yin, Shaohui
; APPLICANT: Cornett, Catherine A.G.
; TITLE OF INVENTION: Transcriptional Control Sequences and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,639
; FILING DATE: 18-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.

```

REGISTRATION NUMBER: 33878  
REFERENCE/DOCKET NUMBER: 69-94  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4254 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Nicotiana tabacum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(1217..1327, 1455..1718, 1806..2182,  
LOCATION: 2259  
LOCATION: ..2477, 2609..2747, 2903..3148, 3262..3558)  
US-08-443-639-7

Query Match 14.3%; Score 32; DB 2; Length 4254;  
Best Local Similarity 48.9%; Pred. No. 1.9;  
Matches 86; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 10 ttataattataattataactgaaccatggtgcccctgctgtttacatgattt 69  
Db 408 TTTATTTTCAAGAAAAAGTTGTCATGTCGCGCGCATGTACACACACACAC 349  
QY 70 ctccgctactattgtatagctgtatataccgtataatgtacatatatttatgaac 129  
Db 348 ATACATTAT 289  
QY 130 atgatcaatgctgtgagttcttcacgtacgttaaggtttcaatgtatgtaagtga 185  
Db 288 ATATATTAATCTATATATTTCTTTTAACTTTTCACAATTTTAGTGTATTTTCTGA 233

RESULT 13  
US-08-446-855A-1/c  
Sequence 1, Application US/08446855A  
Patent No. 5849573  
GENERAL INFORMATION:  
APPLICANT: Stewart, Thomas S  
APPLICANT: Flores, Maria V  
APPLICANT: O'Sullivan, William J  
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl  
TITLE OF INVENTION: phosphate synthetase II  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vanderhye PC  
STREET: 1100 No. 5849573th Glebe Road, 8th Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,855A  
FILING DATE: 06-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mitchard, Leonard C  
REGISTRATION NUMBER: 29,009  
REFERENCE/DOCKET NUMBER: 47-80  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8920 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic  
US-08-446-855A-1

Query Match 14.3%; Score 32; DB 2; Length 8920;  
Best Local Similarity 54.2%; Pred. No. 2.2;  
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 77 actattgtatatacgtgtatataccgtataatgtacatatatttatgaacatgatta 136  
Db 8544 AAT 8485  
QY 137 atgctgtgagttgtcttcacgttaaggtttcaatgtatgtaaggtgcaaac 196  
Db 8484 TACTTTTCATGTTGTTATGGAATGAAAGTTAATAACATACATAAATAAGTAATGTTAAATC 8425

RESULT 14  
US-09-150-741-1/c  
Sequence 1, Application US/09150741  
Patent No. 6183996  
GENERAL INFORMATION:  
APPLICANT: Stewart et al.  
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
Patent No. 6183996  
TITLE OF INVENTION: Synthetase II  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/09/150,741  
CURRENT FILING DATE: 1998-09-10  
EARLIER APPLICATION NUMBER: PL6380  
EARLIER FILING DATE: 1992-12-16  
EARLIER APPLICATION NUMBER: AU93/00617  
EARLIER FILING DATE: 1993-12-02  
EARLIER APPLICATION NUMBER: 08/446,855  
EARLIER FILING DATE: 1995-07-06  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 8920  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
US-09-150-741-1

Query Match 14.3%; Score 32; DB 4; Length 8920;  
Best Local Similarity 54.2%; Pred. No. 2.2;  
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 77 actattgtatatacgtgtatataccgtataatgtacatatatttatgaacatgatta 136  
Db 8544 AAT 8485  
QY 137 atgctgtgagttgtcttcacgttaaggtttcaatgtatgtaaggtgcaaac 196  
Db 8484 TACTTTTCATGTTGTTATGGAATGAAAGTTAATAACATACATAAATAAGTAATGTTAAATC 8425

RESULT 15  
US-08-282-581-4  
Sequence 4, Application US/08282581  
Patent No. 5670349  
GENERAL INFORMATION:  
APPLICANT: Cramer, Carole L.  
APPLICANT: Weissenborn, Deborah L.  
TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM AND



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:04:07 ; Search time 9904.61 Seconds  
(without alignments)  
241.939 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416

Perfect score: 223

Sequence: 1 ctaccataattataattata.....catgaacacccaactegat 223

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estfun:\*\*
  - 2: em\_esthum:\*\*
  - 3: em\_estin:\*\*
  - 4: em\_estom:\*\*
  - 5: em\_estpl:\*\*
  - 6: em\_estba:\*\*
  - 7: em\_estro:\*\*
  - 8: em\_estov:\*\*
  - 9: em\_hic:\*\*
  - 10: gb\_est1:\*\*
  - 11: gb\_est2:\*\*
  - 12: gb\_hic:\*\*
  - 13: gb\_gss:\*\*
  - 14: em\_gss\_fun:\*\*
  - 15: em\_gss\_hum:\*\*
  - 16: em\_gss\_inv:\*\*
  - 17: em\_gss\_pln:\*\*
  - 18: em\_gss\_pro:\*\*
  - 19: em\_gss\_rod:\*\*
  - 20: em\_gss\_vrt:\*\*
  - 21: em\_gss\_other:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45.6	20.4	507	AA605352	AA605352 30388 Lam
C 2	44.4	19.9	368	AU088482	AU088482 AU088482
C 3	43.6	19.6	465	BE948306	BE948306 UI-M-BH3-
C 4	42	18.8	1024	CNS03000	AL222093 Tetraodon
C 5	41.8	18.7	487	AU087926	AU087926 AU087926
C 6	41.8	18.7	500	AU088152	AU088152 AU088152
C 7	41.6	18.7	527	AQ172252	AQ172252 HS_3179_B
C 8	41.6	18.7	599	AQ273686	AQ273686 nbx0030P
C 9	41.6	18.7	711	AZ119444	AZ119444 RPCI-23-1
C 10	41.6	18.7	789	AZ126726	AZ126726 OSJNB007
C 11	40.6	18.2	571	AQ797772	AQ797772 HS_2161_B
C 12	40.6	18.2	605	AZ996260	AZ996260 2M0282F21

C 13	40.4	18.1	583	13	AQ774291	AQ774291 HS_3086_A
C 14	40.4	18.1	595	10	AA394343	AA394343 26024 Lam
C 15	40.4	18.1	845	13	AQ745537	AQ745537 HS_2272_A
C 16	40.2	18.0	525	11	BF460992	BF460992 UI-M-CGOp
C 17	39.8	17.8	612	13	AZ409529	AZ409529 IM0101L12
C 18	39.8	17.8	836	13	AZ908564	AZ908564 RPCI-24-2
C 19	39.6	17.8	456	11	BG836222	BG836222 km06_03b0
C 20	39.4	17.7	458	13	BH040914	BH040914 RPCI-24-2
C 21	39.4	17.7	765	13	AZ705201	AZ705201 RPCI-23-2
C 22	39.4	17.7	939	13	CNS0403P	AL268414 Tetraodon
C 23	39.2	17.6	363	13	AZ698385	AZ698385 RPCI-23-2
C 24	39.2	17.6	1020	13	CNS03AUS	AL235693 Tetraodon
C 25	39.2	17.6	1101	13	CNS00DGO	AL069970 Drosophila
C 26	39	17.5	309	13	AQ542682	AQ542682 RPCI-11-3
C 27	39	17.5	511	11	BG633595	BG633595 GM02540_3
C 28	39	17.5	610	11	BF296565	BF296565 038PBG10
C 29	39	17.5	764	13	AQ915359	AQ915359 nbx0056M
C 30	39	17.5	834	10	AW941790	AW941790 GH07636_3
C 31	38.6	17.3	374	13	BH126294	BH126294 BARC-Sat_
C 32	38.6	17.3	453	13	AQ818698	AQ818698 HS_5268_A
C 33	38.6	17.3	596	13	BH126267	BH126267 BARC-Sat_
C 34	38.4	17.2	500	10	AU087788	AU087788 AU087788
C 35	38.4	17.2	536	13	AZ013131	AZ013131 RPCI-23-3
C 36	38.4	17.2	563	13	AQ561057	AQ561057 HS_5342_B
C 37	38.2	17.1	192	10	AA547848	AA547848 MB3D6V1G0
C 38	38.2	17.1	543	13	AZ060650	AZ060650 RPCI-23-4
C 39	38.2	17.1	604	13	AZ347990	AZ347990 IM0084E22
C 40	38	17.0	514	13	AQ141523	AQ141523 HS_3154_A
C 41	38	17.0	541	13	AZ640886	AZ640886 IM0503E11
C 42	38	17.0	609	13	AQ367613	AQ367613 tox00002B
C 43	37.8	17.0	324	13	BH126635	BH126635 BARC-Satt
C 44	37.8	17.0	412	10	AI991410	AI991410 ws10a07.x
C 45	37.8	17.0	416	13	AZ869448	AZ869448 2M0181N19

ALIGNMENTS

RESULT 1  
AA605352/c  
LOCUS AA605352 507 bp mRNA EST 29-SEP-1997  
DEFINITION 30388 Lambda-PRL2 Arabidopsis thaliana cdna clone 269H8T7, mRNA sequence.  
ACCESSION AA605352  
VERSION AA605352.1 GI:2445880  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 507)  
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,J.L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel,E. and Somerville,C.  
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cdna clones Plant Physiol. 106, 1241-1255 (1994)  
JOURNAL MEDLINE 95148729  
COMMENT Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, Mi  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313tcn@ibm.cl.msu.edu  
Seq primer: T7 dye primer.  
FEATURES  
source Location/Qualifiers  
i . 507  
/organism="Arabidopsis thaliana"  
/strain="var columbia"  
/db\_xref="taxon:3702"

```

/clone="269H87"
/clone_lib="Lambda Zip-Lox; Site.1: Sal; Site.2: Not;
/Note=Vector: Lambda Zip-Lox is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's Lambda Zip-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA.
BASE COUNT      138 a      112 g      144 t      19 others
ORIGIN

```

```

Query Match      20.4%; Score 45.6; DB 10; Length 507;
Best Local Similarity 60.5%; Pred. NO. 0.23;
Matches 75; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 93 tatatatccgtatattatgacatgattatgctgtgagttgtt 152
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 336 TATATATATATGATGATGCCACGATGTATATAACATGACAACTCAATGCTGTGTC 277
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 153 ctcatccgttaaggttccaatgttaattggtgaagtcacaaacccaataatcatgaacac 212
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 276 ATCTCCATAGAGTTGCCAATATGTCATGTTGTAACAGTCAAAACCAAAATCATGAACAC 217
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 213 ccaa 216
|||
DB 216 CAAA 213

```

```

RESULT 2
AU088482/c
LOCUS      AU088482      368 bp      mRNA      EST      27-JAN-2001
DEFINITION AU088482 Sugano Malaria cDNA library Plasmodium falciparum cDNA
clone XPfn7177, mRNA sequence.
ACCESSION AU088482
VERSION    AU088482.1 GI:12390623
KEYWORDS   EST.
SOURCE     Malaria parasite P. falciparum.
ORGANISM   Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE  1 (bases 1 to 368)
AUTHORS    Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE      FULL-malaria: a database for a full-length enriched cDNA library
            from human malaria parasite, Plasmodium falciparum
JOURNAL    Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE    20574754
COMMENT     Contact: Junichi Watanabe
            Institute of Medical Science
            The University of Tokyo, Department of Parasitology
            4-6-1 Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5378
            Fax: 81-3-5449-5410
            Email: jwatanab@manage.ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
            Location/Qualifiers
            1..368
               /organism="Plasmodium falciparum"
               /strain="3b7"
               /db_xref="taxon:5833"
               /clone="XPfn7177"
               /dev_stage="erythrocytic stage"
               /dev_stage="erythrocytic stage"
BASE COUNT      136 a      28 c      37 g      152 t      15 others
ORIGIN

```

```

Query Match      19.9%; Score 44.4; DB 10; Length 368;

```

```

Best Local Similarity 53.2%; Pred. NO. 0.44;
Matches 115; Conservative 0; Mismatches 98; Indels 3; Gaps 1;

QY 2 taccataattataattataattataataactgaacaccatggtgccccctgctgctttac 61
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 240 TACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 181
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 62 atggatttcctccgtactattgtatcactgtatataaccgtata---atgacatata 118
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 180 TTTTATAGAAATGATATATATTTTATATATGATTAATATATATATATATATATA 121
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 119 ttatatgaacatgattatgctgtgagttgtctctcatccgtgaagtttcaatatgta 178
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 120 TTTTAAAAAATTTTAAATTTTATCTGTTTCGTACCCCTATATCTATAAATAT 61
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 179 atgggtgaagagtcacaaacccaataatcatgaacccc 214
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 60 TTCAATAGATATATAAACNCAANATCATATATATCC 25
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

RESULT 3
BE948306
LOCUS      BE948306      465 bp      mRNA      EST      03-OCT-2000
DEFINITION UI-M-BH3-awp-b-11-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-awp-b-11-0-UI 3', mRNA sequence.
ACCESSION BE948306
VERSION    BE948306.1 GI:10526065
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 465)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    9704477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mestr@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            cerebellum tissue cDNA Library Preparation: M.B. Soares Lab Clone
            distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
            GENETICS. It should be noted that Bento Soares is generating a
            small number of additional specialized non-redundant arrays of BMAP
            cDNAs whose availability will be considered under appropriate and
            limited collaborative arrangements The following repetitive
            elements were found in this cDNA sequence: 255-326, >Lx98LINE/L1
            Seg primer: M13 Forward
            POLYA-yes.

```

```

FEATURES
            Location/Qualifiers
            1..465
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="UI-M-BH3-awp-b-11-0-UI"
               /dev_stage="27-32 days"
               /dev_stage="27-32 days"
               /lab_host="DHI0B (Life Technologies)"
               /note="Vector: pMT30-Pac (Pharmacia) with a modified
               polylinker; Site.1: Not I; Site.2: Eco RI; The
               NIH_BMAP_M_S4 library is a subtracted library of a series,
               ultimately derived from a mixture of individually tagged
               normalized libraries from ten regions of the mouse brain

```



BASE COUNT	213 a	34 C	23 g	211 t	6 others
ORIGIN					

Query Match 18.7%; Score 41.8; DB 10; Length 487;  
Best Local Similarity 52.4%; Pred. NO. 1.9;  
Matches 88; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

**Oy** 8 aattataattataattatataatactgaacaccatggtgccccctgcgtgctttacatgat 67  
|| ||| || | ||| || | ||| || | ||| || | ||| || | ||| ||  
**Dd** 168 AAATATATTGAAATATATTATGAATATATATATGACTATTTTCTTCATGCTGAT 109

[illegible]

DD TGTATTANCTATATATATATATATATATATGTTAAATAAATAATAGTAGAAC 49  
QY 128 acatgattaatgcttgtaggttgttctcatccgtaagaggtttcgaatat 175

Db  
48 GATTTTCTTTTCTTTTATTTTATATATATATAAATCCGNAAAAATTTTNCAAAAT 1

RESULT	6
AU088152	
LOCUS	
DEFINITION	500 bp mRNA EST 27-JAN-2001
	AU088152 Sugano Malaria CDNA library Plasmodium falciparum cDNA
	clone xPFn606 similar to Plasmodium falciparum microsatellite TA61
	sequence, mRNA sequence.

ACCESSION	AU088152
VERSION	AU088152.1
KEYWORDS	GI:12390293
SOURCE	EST.
ORGANISM	malaria parasite <i>P. falciparum</i> . Plasmodium falciparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 500)
AUTHORS	Watanabe, J., Sasaki, M., Suzuki, Y. and Sugano, S.
TITLE	FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum
JOURNAL	Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE	20574754

CONTACT: Junichi Watanabe  
Institute of Medical Science  
The University of Tokyo, Department of Parasitology  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Tel: 81-3-5449-5378  
Fax: 81-3-5449-5410  
Email: jwatanab@manage.ims.u-tokyo.ac.jp  
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156(1997).

BASE COUNT	224 a	37 c	38 g	182 t	19 others
FEATURES					
source					
Location/Qualifiers					
1. .500					
/organism="Plasmodium falciparum"					
/strain="3D7"					
/db_xref="taxon:5833"					
/clone="XPFn6606"					
/clone_lib="Sugano Malaria cDNA library"					
/dev_stage="erythrocytic stage"					

```

Query Match      18.7%; Score 41.8; DB 10; Length 500;
Best Local Similarity 55.5%; Pred. No. 1.9;
Matches 76; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```

83 *tgtatacgtgtatatatacogtataatgtacatatatttatgaacatgattaatgcgtt* 142

63 TATTTATNCTAAATATATGATTACATTATATATATATATATATATATATATGTTTA 122

143 GTGAGTTGTTCTCATCCGTAAGAGTTTCAATATGTAATGTTGAAGAGTCAAAACCCAAAA 202

• QY 143 gtgagttgttctcatccgtaagaggtttcaatatgtaatggtgaagagtcacaaacccaaaa 202

Db 123 TTNATTTATTAAATATATAATATATATATATATATGGAAGAANGAAAGAAAA 182

Qy 203 tcatgaacaccccaact 219

Db 183 AAATTTAAAGCAAATT 199

RESULT	7
AQ172252/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 527)  
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and  
 Hood, L.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahabir G. Wallace to: ucd@ucl.ac.uk

High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Sequence Tagged Connector  
 Plate: 3179 row: D column: 22  
 Class: BAC ends  
 High quality sequence stop: 527.

```

FEATURES
  source
    Location/Qualifiers
      1..527
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_plate=J179 Col=22 Row=D"
        /clone_lib="Cir Approved Human Genomic Sperm Library D"
        /sex="male"
        /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
          E-Coli DH10B"
      195 a 56 c 55 g 218 t 3 others
BASE COUNT
ORIGIN

```

Query Match 18.7%; Score 41.6; DB 13; Length 527;  
Best Local Similarity 70.0%; Pred. NO. 2.1;  
Matches 56; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

57 tttaacatggattccgcgtactatttgatatcagcggtatatataaccgtataatgtacata 116  
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
403 TATATAGATATATAAAGATATAAATTATATATATTTTATATATAAAAATTAATGATATATA 344

117 tatttatatgaacatgatta 136  
||||||| | | | |  
343 TATTTATATATAAAATATA 324

RESULT	8
LOCUS	AQ273686 599 bp DNA
DEFINITION	nxbb0030P21f CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION	nxbb0030P21f, DNA sequence.
VERSION	AQ273686
KEYWORDS	AQ273686.1 GI:3827001 GSS.





ACCESSION AZ126726  
 VERSION AZ126726.1 GI:8201718  
 KEYWORDS GSS.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 789)  
 REFERENCE Wing, R.A. and Dean, R.A.  
 A BAC End Sequencing Framework to Sequence the Rice Genome Unpublished (1998)  
 JOURNAL Contact: Wing RA  
 COMMENT Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: GTAAACGACGCCAGTG  
 Class: BAC ends  
 High quality sequence start: 21  
 High quality sequence stop: 683.  
 Location/Qualifiers  
 1. 789  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="OSJNB0079C19f"  
 /clone\_lib="CUGI Rice BAC Library (ECORI)"  
 /tissue\_type="Leaf"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pBACindigo; Site\_1: EcoRI; Site\_2: EcoRI;  
 Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate.  
 Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."  
 BASE COUNT 292 a 124 c 111 g 253 t 9 others  
 ORIGIN

Query Match 18.7%; Score 41.6; DB 13; Length 789;  
 Best Local Similarity 50.9%; Pred. No. 2.1;  
 Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
 QY 9 attataattataataactgaaccatgggtgccccgcgttcattacatgatt 68  
 Db 745 ATTATATATTATATATATAAATATATATTTTATATTCATTTTTTNTN 686  
 QY 69 tctccgctactattgtatataacgtatataacgtatataacgtatatttatgaa 128  
 Db 685 NATANNNTATTTTATATATATATATATATATATATATATATATATATA 626  
 QY 129 catgattaaagctgtgagtggtcttcacccgtaagagtttcaatgt 177  
 Db 625 TATATATCTTCATTCAGCATGTTTCACATACAGATAGCTTTTAGGTAT 577

RESULT 11  
 A0797772/c  
 LOCUS 571 bp DNA GSS 09-AUG-1999  
 DEFINITION HS\_2161\_B2\_E10\_77C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2161 Col=20 Row=J, DNA sequence.  
 ACCESSION A0797772  
 VERSION A0797772.1 GI:5715030  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 571)  
 REFERENCE Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and Hood, L.  
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).  
 BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 2161 row: J column: 20  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 571.  
 Location/Qualifiers  
 1. 571  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=2161 Col=20 Row=J"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"  
 BASE COUNT 197 a 48 c 83 g 241 t 2 others  
 ORIGIN  
 Query Match 18.2%; Score 40.6; DB 13; Length 571;  
 Best Local Similarity 58.8%; Pred. No. 3.7;  
 Matches 70; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 60 acatgatttctccgtactattgttatcgtgtatataacgtatataacgtatataatgtacatatat 119  
 Db 192 ACATATATATACAT 133  
 QY 120 ttatatgaacatgattaatgctgtgagttcttcctccgtaagagtttcaatgta 178  
 Db 132 ATATATACAT 74  
 RESULT 12  
 A2996260/c  
 LOCUS 605 bp DNA GSS 27-APR-2001  
 DEFINITION 2M0282F21F Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0282F21 F, DNA sequence.  
 ACCESSION A2996260  
 VERSION A2996260.1 GI:13867487  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:47:43 ; Search time 2553.1 Seconds  
(without alignments)  
1440.943 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416  
Perfect score: 223  
Sequence: 1 ctaccataataataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sv.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	223	100.0	416	6	AX127757	AX127757 Sequence
c	189.8	85.1	415	6	AX127755	AX127755 Sequence
3	47.8	21.4	186185	9	AC008817	AC008817 Homo sapi
4	46.4	20.8	110434	9	HS066722	U66722 Human chrom
5	46	20.6	138350	2	AC010801	AC010801 Homo sapi
6	46	20.6	142126	2	AL596266	AL596266 Homo sapi
7	45.6	20.4	81503	2	AC091684	AC091684 Homo sapi
c	45.2	20.3	146349	2	AC007863	AC007863 Trypanoso
8	45.2	20.3	149241	2	AC068887	AC068887 Homo sapi
9	45.2	19.8	158236	9	AC011631	AC011631 Homo sapi
10	44.2	19.8	216431	2	AL354771	AL354771 Homo sapi
11	44.2	19.7	40917	6	AX033911	AX033911 Sequence
12	44	19.7	41008	6	AX033912	AX033912 Sequence
13	44	19.7	80155	9	AF042484	AF042484 Homo sapi
14	44	19.7	80155	9	AF222855	AF222855 Homo sapi
15	44	19.7	80622	9	AF222855	AF222855 Homo sapi
c	43.8	19.6	175959	2	AL450429	AL450429 Homo sapi
16	43.6	19.6	228	8	AY020821	AY020821 Oryza sat
17	43.6	19.6	234	8	AY018525	AY018525 Oryza sat
18	43.6	19.6	185499	2	AC090578	AC090578 Homo sapi
c	43.6	19.6	187967	2	AC026089	AC026089 Homo sapi
20	43.6	19.5	145358	2	AC091086	AC091086 Oryza sat
c	43.4	19.4	108523	9	AC000159	AC000159 Homo sapi
22	43.2	19.4	135237	2	AP000814	AP000814 Homo sapi
c	43.2	19.4	156736	9	AC006502	AC006502 Homo sapi
23	43.2	19.3	191962	2	AC048342	AC048342 Homo sapi
25	43	19.3	193962	2	AC068503	AC068503 Homo sapi
26	43	19.3	203718	9	AC009486	AC009486 Homo sapi
27	43	19.2	156060	2	AC004153	AC004153 Plasmodiu
c	42.8	19.2	165114	2	AC011854	AC011854 Homo sapi
28	42.8	19.2	175020	9	AC022317	AC022317 Homo sapi
c	42.8	19.2	185702	2	AP002963	AP002963 Homo sapi
31	42.8	19.2	211316	9	AC009294	AC009294 Homo sapi
c	42.8	19.2	269661	2	AC023911	AC023911 Homo sapi
33	42.8	19.0	69058	9	AF222856	AF222856 Homo sapi
34	42.4	19.0	80202	9	AF222854	AF222854 Homo sapi
35	42.4	19.0	101584	9	CNS01DS5	AL121655 BAC seque
c	42.4	19.0	155390	2	AL596118	AL596118 Homo sapi
37	42.4	19.0	162958	2	AC011022	AC011022 Homo sapi
c	42.4	19.0	168519	2	AL355340	AL355340 Homo sapi
39	42.4	19.0	168671	2	AC019247	AC019247 Homo sapi
40	42.4	19.0	173435	2	AC015715	AC015715 Homo sapi
c	42.4	19.0	176432	2	AL391648	AL391648 Homo sapi
42	42.4	19.0	180942	33	AL512845	AL512845 Human DNA
c	42.4	18.9	66823	2	AC073039	AC073039 Homo sapi
44	42.2	18.9	110880	9	AC004878	AC004878 Homo sapi
c	42.2	18.9	110880	9	AC004878	AC004878 Homo sapi

ALIGNMENTS

RESULT 1	AX127757	416 bp	DNA	PAT	15-MAY-2001
LOCUS	Sequence 10 from Patent WO0131042.				
DEFINITION	AX127757				
ACCESSION	AX127757.1	GI:14134404			
VERSION					
KEYWORDS	synthetic construct.				
SOURCE	synthetic construct				
ORGANISM	artificial sequence.				
REFERENCE	1 (bases 1 to 416)				
AUTHORS	Weston,B. and de Beuckeleer,M.				
TITLE	Male-sterile brassica plants and methods for producing same				
JOURNAL	Patent: WO 0131042-A 10 03-MAY-2001;				
FEATURES	Aventis CropScience N.V. (BE)				
source	Location/Qualifiers				
	1. .416				
	/organism="synthetic construct"				
	/db_xref="taxon:32630"				
	/note="3' border flanking region of elite event MS-B2"				

PRI 11-MAR-1997  
3B sequence.

ACCESSION U66722  
VERSION U66722.1 GI:1871363  
KEYWORDS human,  
SOURCE Homo sapiens  
ORGANISM  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 110434)  
Bolog, F., Gemmill, R.M., West, J., Robinson, M., Robinson, L., Li, E.,  
Roche, J., Todd, S., Waggoner, B., Lundstrom, R., Jacobson, J.,  
Mullokandov, M.R., Klinger, H. and Drabkin, H.A.  
Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B  
Hum. Mol. Genet. 6 (2), 193-203 (1997)  
TITLE 9721778  
JOURNAL 2 (bases 1 to 110434)  
MEDLINE Drabkin, H.A., Bolog, F., West, J., Roche, J., Robinson, L.,  
AUTHORS Robinson, M., Lundstrom, R., Li, E.F., Klinger, H. and Gemmill, R.M.  
Direct Submission  
TITLE Submitted (12-AUG-1996) Medical Oncology, University of Colorado  
JOURNAL Health Sciences Center, 4200 E 9th Ave Box B171, Denver, CO 80262,  
USA  
COMMENT On Mar 8, 1997 this sequence version replaced gi:1669854.  
FEATURES  
source 1. 110434  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/map="3p14.2"  
325. 828  
/rpt\_family="MER6"  
repeat\_region 920. 1040  
/rpt\_family="microsatellite"  
933. 991  
/rpt\_family="HSAT1"  
1287. 1579  
/rpt\_family="ALU"  
repeat\_region complement(4624. 4730)  
/rpt\_family="MIR"  
7863. 7887  
LTR /note="LTR8"  
repeat\_region 8276. 8340  
/rpt\_family="microsatellite"  
LTR complement(9352. 9405)  
/note="LTR5"  
10859. 12771  
/rpt\_family="Tigger1"  
repeat\_region complement(13165. 13731)  
/rpt\_family="L1"  
13729. 16089  
repeat\_region /rpt\_family="L1"  
complement(17236. 17352)  
repeat\_region /rpt\_family="ALU"  
18542. 18838  
repeat\_region /rpt\_family="ALU"  
complement(20464. 20563)  
repeat\_region /rpt\_family="MIR"  
complement(21303. 21595)  
repeat\_region /rpt\_family="ALU"  
23140. 23157  
repeat\_region /rpt\_family="HSAT1"  
24152. 24196  
repeat\_region /rpt\_family="MLT1D"  
24219. 24274  
repeat\_region /rpt\_family="MLT1D"  
25663. 25913  
repeat\_region /rpt\_family="ALU"  
26593. 26860  
repeat\_region /rpt\_family="MER7"  
26862. 26936  
repeat\_region /rpt\_family="MER7"  
complement(28463. 28533)  
repeat\_region /rpt\_family="MER5"  
complement(28903. 29000)  
repeat\_region  
/rpt\_family="MIR"  
32445. 32515  
/note="LTR3"  
33949. 34268  
/rpt\_family="L1"  
repeat\_region 34584. 35022  
/rpt\_family="L1"  
35062. 35099  
/note="LTR6"  
complement(35225. 35379)  
/rpt\_family="MIR"  
35771. 35799  
/rpt\_family="ALR"  
36009. 36120  
/note="polypurine"  
complement(36138. 36295)  
/rpt\_family="ALU"  
repeat\_region complement(37577. 37610)  
/rpt\_family="XTR"  
39771. 39912  
/rpt\_family="microsatellite"  
43071. 43274  
/rpt\_family="microsatellite"  
complement(43199. 43410)  
/rpt\_family="ALU"  
repeat\_region complement(47356. 47550)  
/rpt\_family="MER20"  
complement(47798. 47908)  
/rpt\_family="TARI"  
repeat\_region complement(48812. 48924)  
/rpt\_family="MER5"  
50409. 50541  
/rpt\_family="MSTB"  
50444. 50541  
/rpt\_family="MLT1D"  
complement(52329. 52447)  
/rpt\_family="MIR"  
repeat\_region complement(53748. 54028)  
/rpt\_family="ALU"  
complement(56825. 57121)  
/rpt\_family="ALU"  
58470. 58533  
repeat\_region /rpt\_family="MIR"  
61841. 62112  
repeat\_region /rpt\_family="ALU"  
complement(62331. 62604)  
/rpt\_family="ALU"  
64061. 64195  
repeat\_region /rpt\_family="L1"  
64331. 64468  
/note="similar to EST R787000"  
complement(65817. 66065)  
/rpt\_family="L1"  
66100. 66588  
repeat\_region /rpt\_family="MLT1D"  
66784. 67036  
repeat\_region /rpt\_family="L1"  
67056. 67843  
repeat\_region /rpt\_family="L1"  
complement(68636. 68913)  
/rpt\_family="ALU"  
69074. 69099  
/note="LTR6"  
69253. 69371  
repeat\_region /rpt\_family="ALU"  
71295. 71324  
repeat\_region /rpt\_family="XBR"  
complement(72381. 72710)  
/rpt\_family="ALU"  
repeat\_region complement(72953. 73047)  
/rpt\_family="L1"  
complement(74142. 74787)  
/rpt\_family="L1"

```

repeat_region complement(74996..75295)
/rpt_family="ALU"
repeat_region complement(75316..75446)
/rpt_family="L1"
misc_feature 75959..76959
/note="CC19 telomeric breakpoint"
LTR 78426..78469
/note="LTR1"
repeat_region 81581..81638
/rpt_family="MIR"
repeat_region complement(82365..82488)
/rpt_family="ALU"
repeat_region complement(83558..83843)
/rpt_family="ALU"
repeat_region 85341..85493
/rpt_family="MER42"
misc_feature 85859..85960
/note="HPV viral intergration site"
LTR 87350..87452
/note="LTR1"
repeat_region 88396..88683
/rpt_family="ALU"
repeat_region 90782..91043
/rpt_family="L1"
repeat_region 96484..96611
/rpt_family="HSAT1"
repeat_region 96613..96910
/rpt_family="ALU"
repeat_region 99040..99081
/rpt_family="HSAT1"
repeat_region 99834..100410
/rpt_family="ALU"
repeat_region 101347..101767
/rpt_family="THE1"
repeat_region complement(108148..108193)
/rpt_family="L1"
repeat_region complement(108781..110294)
/rpt_family="L1"

BASE COUNT 34970 a 22068 c 20889 g 32500 t 7 others
ORIGIN

Query Match 20.8%; Score 46.4; DB 9; Length 110434;
Best Local Similarity 53.3%; Pred. No. 0.13;
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1 ctaccataattataataattataataactgaaccatggtgccccctgcttta 60
Db 42950 CTGCCCTAAATATACATATACATATAATATGTATATGTGTGTGTATATA 43009

QY 61 catggatttcgcgtactattgtatcgtgtatataccgtataatgtacatatatt 120
Db 43010 TATGGACACACACATATATATATGTGTGTATATGTGCATATGTATATA 43069

QY 121 tatatgaacatgatcaatgctgtgagttgttctcatccgaagatttcaatgtaatt 180
Db 43070 TATACGCACATACGTATGTGTATATACGTGTAGATATATGTGTATATATGTGTG 43129

QY 181 gggtg 184
Db 43130 TGTG 43133

RESULT 5
LOCUS AC010801 138350 bp DNA HTG 30-MAR-2000
DEFINITION Homo sapiens clone RP11-2L13, WORKING DRAFT SEQUENCE, 19 unordered
pieces.
ACCESSION AC010801
VERSION AC010801.3 GI:7341823
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 138350)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens, clone RP11-2L13  
 Unpublished

2 (bases 1 to 138350)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Balgwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeAtellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, P., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 30, 2000 this sequence version replaced gi:6139142.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
 Center project name: L2660  
 Center clone name: 2.L13  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 125609 bases at least Q40  
 Consensus quality: 132073 bases at least Q30  
 Consensus quality: 134441 bases at least Q20  
 Insert size: 139000; agarose-fp  
 Insert size: 136550; sum-of-contigs  
 Quality coverage: 3.6 in Q20 bases; agarose-fp  
 Quality coverage: 3.6 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 19 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1169: contig of 1169 bp in length  
 1170 1269: gap of 100 bp  
 1270 4174: contig of 2905 bp in length  
 4175 4274: gap of 100 bp  
 4275 6879: contig of 2605 bp in length  
 6880 6979: gap of 100 bp  
 6980 10304: contig of 3325 bp in length  
 10305 10404: gap of 100 bp  
 10405 13133: contig of 2729 bp in length  
 13134 13233: gap of 100 bp  
 13234 16716: contig of 3483 bp in length  
 16717 16816: gap of 100 bp  
 16817 21929: contig of 5113 bp in length  
 21930 22029: gap of 100 bp  
 22030 26268: contig of 4239 bp in length  
 26269 26368: gap of 100 bp  
 26369 31448: contig of 5080 bp in length

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 COMMENT





```

/note="assembly_fragment:01221
fragment_chain:1
clone_end:SP6
vector_side:left"
misc_feature
6553..142126
/note="assembly_fragment:01911
fragment_chain:1
clone_end:T7
vector_side:right"
BASE COUNT 47396 a 26793 c 24991 g 42846 t 100 others
ORIGIN

Query Match      20.6%; Score 46; DB 2; Length 142126;
Best Local Similarity 54.0%; Pred. No. 0.16;
Matches 94; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 10 ttataataattataataactgaacacatggctgcccctgctgtttacatgattt 69
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124601 TTATGCCATACACTTAGATATATATATATCTCATAGAGATACAGATTATATATATATAT 124660
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 70 ctccgctactattgtatcacgtgtatatataaccgtataatgaatgacatatatttatgaac 129
      | |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124661 ATATATATATATATATATATATATATATATACACTCATATGTTGATATATTTTAAAGGTATCC 124720
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 130 atgattaatgcttgaggtgttctcaccgtgaagaggttccaatgtaagtgt 183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124721 CTTCTTAATACAGTGCATGACACACCTATATCTATACCTATATATCTATATCTACAGTT 124774

RESULT 7
AC091684
LOCUS      AC091684      81503 bp      DNA      HTG      17-MAY-2001
DEFINITION Homo sapiens chromosome UNK clone RP13-581E15, *** SEQUENCING IN
            PROGRESS ***, 41 unordered pieces.
ACCESSION AC091684
VERSION   AC091684.1 GI:14140336
KEYWORDS  HTG; HTGS_PHASE1.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS   Waterston,R.H.
TITLE     The sequence of Homo sapiens clone
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 81503)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (17-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1407 1508: contig of 1406 bp in length
* 1507 3416: contig of unknown length
* 3417 3516: gap of unknown length
* 3517 6161: contig of 2645 bp in length
* 6162 6261: gap of unknown length
* 6262 7972: contig of 1711 bp in length
*
* 8072: gap of unknown length
* 11094: contig of 3022 bp in length
* 11194: gap of unknown length
* 13957: contig of 2763 bp in length
* 14057: gap of unknown length
* 16589: contig of 2532 bp in length
* 16690: gap of unknown length
* 18911: contig of 2222 bp in length
* 19011: gap of unknown length
* 21691: contig of 2680 bp in length
* 21791: gap of unknown length
* 24079: contig of 2288 bp in length
* 24180: gap of unknown length
* 26194: contig of 2015 bp in length
* 26294: gap of unknown length
* 29462: contig of 2868 bp in length
* 29462: gap of unknown length
* 33740: contig of 4478 bp in length
* 33840: gap of unknown length
* 35244: contig of 1404 bp in length
* 35345: gap of unknown length
* 36397: contig of 1053 bp in length
* 36497: gap of unknown length
* 37724: contig of 1227 bp in length
* 37824: gap of unknown length
* 39046: contig of 1222 bp in length
* 39146: gap of unknown length
* 40876: contig of 1730 bp in length
* 40976: gap of unknown length
* 42169: contig of 1193 bp in length
* 42269: gap of unknown length
* 43531: contig of 1262 bp in length
* 43631: gap of unknown length
* 44992: contig of 1361 bp in length
* 45092: gap of unknown length
* 46384: contig of 1292 bp in length
* 46484: gap of unknown length
* 47767: contig of 1283 bp in length
* 47867: gap of unknown length
* 49395: contig of 1528 bp in length
* 49495: gap of unknown length
* 51364: contig of 1869 bp in length
* 51464: gap of unknown length
* 53294: contig of 1830 bp in length
* 53394: gap of unknown length
* 54985: contig of 1591 bp in length
* 55085: gap of unknown length
* 56463: contig of 1378 bp in length
* 56563: gap of unknown length
* 58670: contig of 2107 bp in length
* 58770: gap of unknown length
* 60099: contig of 1329 bp in length
* 60199: gap of unknown length
* 61632: contig of 1433 bp in length
* 61732: gap of unknown length
* 63819: contig of 2087 bp in length
* 63919: gap of unknown length
* 65179: contig of 1260 bp in length
* 65279: gap of unknown length
* 67578: contig of 2299 bp in length
* 67678: gap of unknown length
* 70151: contig of 2473 bp in length
* 70251: gap of unknown length
* 72332: contig of 1981 bp in length
* 72332: gap of unknown length
* 74165: contig of 1833 bp in length
* 74265: gap of unknown length
* 75644: contig of 1379 bp in length
* 75744: gap of unknown length
* 77594: contig of 1850 bp in length
* 77694: gap of unknown length
* 78888: contig of 1194 bp in length
* 78988: gap of unknown length

```

[illegible]







Query Match	19.7%	Score 44	DB 6	Length 40917
Best Local Similarity	58.3%	Pred. No. 0.52		
Matches	77	Conservative	0	Mismatches 55; Indels 0; Gaps 0
QY 5	cataattataattataataataataactgaacacatggtgccccctgctgctttacatg 64			
Db 15579	CATAAATATATACATAAATATGCTATATATGCTATATAGACATAAATATGCTATATATG 15638			
QY 65	gattctccgctactatttgtatcacgtgtgtatatataaccgtataatgtacatatattata 124			
Db 15639	TATATATACATAAATATGCTATATATGCTATATAGACATAAATATGCTATATATGCTATA 15698			
QY 125	tgaacatgatta 136			
Db 15699	TAGACATAAATA 15710			
RESULT 13				
AX033912	LOCUS AX033912 41008 bp DNA PAT 21-SEP-2000			
DEFINITION	Sequence 4 from Patent WO9851790.			
ACCESSION	AX033912			
VERSION	AX033912.1 GI:10280480			
KEYWORDS	unidentified.			
SOURCE	unidentified			
ORGANISM	unclassified.			
REFERENCE	1 (bases 1 to 41008)			
AUTHORS	Cancilla,M.R., Choo,K.H. and Du,S.D.			
TITLE	A novel nucleic acid molecule			
JOURNAL	Patent: WO 9851790-A 4 19-NOV-1998;			
FEATURES	CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)			
source	Location/Qualifiers			
	1..41008			
	/organism="unidentified"			
	/db_xref="taxon:32644"			
BASE COUNT	11778 a 8413 c 8415 g 12398 t 4 others			
ORIGIN				
Query Match	19.7%	Score 44	DB 6	Length 41008
Best Local Similarity	58.3%	Pred. No. 0.52		
Matches	77	Conservative	0	Mismatches 55; Indels 0; Gaps 0
QY 5	cataattataattataataataactgaacacatggtgccccctgctgctttacatg 64			
Db 15314	CATAAATATATACATAAATATGCTATATATGCTATATAGACATAAATATGCTATATATG 15373			
QY 65	gattctccgctactatttgtatcacgtgtgtatatataaccgtataatgtacatatattata 124			
Db 15374	TATATATACATAAATATGCTATATATGCTATATAGACATAAATATGCTATATATGCTATA 15433			
QY 125	tgaacatgatta 136			
Db 15434	TAGACATAAATA 15445			
RESULT 14				
AF042484	LOCUS AF042484 80155 bp DNA PRI 22-JUN-2000			
DEFINITION	Homo sapiens sequence of the 80 kb core centromere protein binding region of the mar del(10) neocentromere corresponding to human chromosome 10q25.2, complete sequence.			
ACCESSION	AF042484			
VERSION	AF042484.1 GI:4205782			
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 80155)			
	Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.A.			

```

TITLE      Sequence analysis of an 80 kb human neocentromere
JOURNAL    Hum. Mol. Genet. 8 (2), 217-227 (1999)
MEDLINE    99135901
PUBMED     9931329
REFERENCE  2 (bases 1 to 80155)
AUTHORS    Barry,A.E., Howman,E.V., Cancelli,M.R., Saffery,R. and Choo,A.
TITLE      Direct Submission
JOURNAL    Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch
           Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,
           Parkville, Melbourne, Victoria 3052, Australia
FEATURES   Location/Qualifiers
            source          1..80155
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /chromosome="10"
                           /map="10q25.2"
                           /note="Sequence information obtained from BAC clones
                           prepared from cell line BE2cl-18-5f"
repeat_region complement(join(49..525,3628..4545,4849..5053))
                           /rpt_family="L1MEc"
repeat_region join(560..597,3184..3281)
                           /rpt_type="dispersed"
                           /rpt_family="MER58"
repeat_region complement(join(598..2553,2716..3183))
                           /rpt_family="RIGGERI"
                           /rpt_type="dispersed"
                           /standard_name="T1"
                           /note="100% homology"
                           /rpt_type="direct"
                           /rpt_unit="(CAGGCACAGTGG)2
                           2570..2715"
repeat_region /rpt_family="AluSg/x"
                           /rpt_type="dispersed"
                           3021..3050
                           /standard_name="T2"
                           /note="83.3% homology"
                           /rpt_type="direct"
                           /rpt_unit="(TAACAAAGTG)3
                           3283..3358"
repeat_region /rpt_family="AluSx"
                           /rpt_type="dispersed"
                           4173..4211
                           /standard_name="T3"
                           /note="82.1% homology"
                           /rpt_type="direct"
                           /rpt_unit="(AAAAATAATTT)3
                           complement(4546..4848)
                           /rpt_family="AluSx"
                           /rpt_type="dispersed"
                           complement(5909..6208)
                           /rpt_family="AluSg"
repeat_region join(6232..6573,6638..6688)
                           /rpt_type="dispersed"
                           /rpt_family="MLT2CB"
                           /rpt_type="dispersed"
                           6598..6645
                           /standard_name="AC24"
                           /note="100% homology"
                           /rpt_type="tandem"
                           /rpt_unit="AC
                           6598..6637
                           /note="microsatellite"
                           /rpt_type="tandem"
                           /rpt_unit="CA
                           7091..7134
                           /rpt_family="MIR"
                           /rpt_type="dispersed"
                           7808..7911
                           /rpt_family="L2"
                           /rpt_type="dispersed"
                           8349..8524
                           /rpt_type="dispersed"
                           /rpt_family="AluSx"
                           /rpt_type="dispersed"
                           complement(8839..9124)
                           /rpt_family="AluJb"
                           /rpt_type="dispersed"
                           9180..9280
                           /rpt_family="MIR"
                           /rpt_type="dispersed"
                           10070..10170
                           /note="low complexity"
                           /rpt_family="purine-rich"
                           10097..10136
                           /standard_name="GAI"
                           /note="82.5% homology"
                           /rpt_type="direct"
                           /rpt_unit="(GAAAGAAAGG)4
                           10175..10339
                           /rpt_family="FRAM"
                           /rpt_type="dispersed"
                           complement(10899..11197)
                           /rpt_family="AluSx"
                           /rpt_type="dispersed"
                           join(11627..11757,12434..12512,13443..13562)
                           /note="598% identity"
                           /db_xref="dbEST:AA774571"
                           11996..12028
                           /standard_name="AT4"
                           /note="100% homology"
                           /rpt_type="tandem"
                           /rpt_unit="(ATT)8
                           complement(12010..12312)
                           /rpt_family="AluJb"
                           /rpt_type="dispersed"
                           12327..12433
                           /rpt_family="L2"
                           /rpt_type="dispersed"
                           complement(13946..14235)
                           /rpt_family="AluSg"
                           /rpt_type="dispersed"
                           15130..15159
                           /note="low complexity"
                           /rpt_family="AT-rich"
                           15164..15753
                           /standard_name="AT28"
                           /note="82.1% homology"
                           /rpt_type="direct"
                           /rpt_unit="(ATGTATATATGTGTATATAGACATAAT)21
                           join(15170..15289,15410..15612)
                           /note="microsatellite"
                           /rpt_type="tandem"
                           /rpt_unit="TA
                           15617..15731
                           /note="microsatellite"
                           /rpt_type="tandem"
                           /rpt_unit="CATA
                           16004..16100
                           /rpt_family="L2"
                           /rpt_type="dispersed"
                           complement(16128..16364)
                           /rpt_family="L2"
                           /rpt_type="dispersed"
                           complement(16407..16592)
                           /rpt_family="L2"
                           /rpt_type="dispersed"
                           complement(16657..16889)
                           /rpt_family="MIR"
                           /rpt_type="dispersed"
                           16914..16936
                           /note="low complexity"
                           /rpt_family="AT-rich"
                           17102..17386
                           /rpt_family="AluSx"
                           /rpt_type="dispersed"

```





```

/rpt_family="MIR"
repeat_region 34689. .34995
/rpt_family="Alu"
repeat_region complement(36349. .36437)
/rpt_family="MIR"
repeat_region complement(38796. .39079)
/rpt_family="Alu"
repeat_region 39441. .39693
/rpt_family="Alu"
repeat_region complement(43176. .43237)
/rpt_family="MIR"
repeat_region complement(45203. .45478)
/rpt_family="Alu"
repeat_region 45892. .46025
/rpt_family="MIR"
repeat_region complement(46618. .46851)
/rpt_family="Alu"
repeat_region 48623. .48939
/rpt_family="Alu"
repeat_region complement(49048. .50698)
/rpt_family="L1"
repeat_region complement(50710. .50972)
/rpt_family="Alu"
repeat_region complement(51031. .51192)
/rpt_family="L1"
repeat_region complement(51450. .51739)
/rpt_family="Alu"
repeat_region complement(53270. .53562)
/rpt_family="Alu"
repeat_region 54261. .54808
/rpt_family="L1"
repeat_region 59460. .60034
/rpt_family="L1"
repeat_region 60349. .60633
/rpt_family="Alu"
repeat_region complement(60669. .60801)
/rpt_family="MIR"
repeat_region complement(61684. .61852)
/rpt_family="MER5"
repeat_region 63466. .63753
/rpt_family="Alu"
repeat_region 64063. .64228
/rpt_family="MER5"
repeat_region 64446. .64521
/rpt_family="MIR"
repeat_region 65794. .65883
/rpt_family="MER5"
repeat_region 66772. .66951
/rpt_family="MER20"
repeat_region complement(67895. .68416)
/rpt_family="MER41"
repeat_region complement(73192. .73470)
/rpt_family="Alu"
repeat_region 74086. .74367
/rpt_family="Alu"
repeat_region 75916. .76183
/rpt_family="Alu"
repeat_region complement(77771. .78059)
/rpt_family="Alu"
BASE COUNT 23195 a 16614 c 16828 g 23985 t
ORIGIN

```

```

Query Match 19.7% Score 44; DB 9; Length 80622;
Best Local Similarity 58.3%; Pred. No. 0.5;
Matches 77; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Oy 5 cataattataattataattactgaaccatggtgccctgctgttcatcg 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15579 CATAAATATATACATAAATATGTATATGTATATAGACATAAATATGTATATATGTG 15638
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 65 gattctccgtactattgtgtacgtgtatataccgtataatgtacatatattata 124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 15639 TATATATACATAAATATGTATATATGTATATATAGACATAAATATGTATATGTGATA 15698
Oy 125 tgaacatgatta 136
| ||||| ||
Db 15699 TAGACATAAATA 15710

```

Search completed: February 15, 2002, 18:53:51  
Job time: 20470 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:01:55 ; Search time 868.33 Seconds  
(without alignments)  
220.174 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416

Perfect score: 223  
Sequence: 1 ctaccataattataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	223	100.0	416	22	Left (3') border f
C	189.8	85.1	415	22	Right (5') border
3	50.4	22.6	244	22	Oligonucleotide D1
4	50	22.4	936	22	Oligonucleotide D1
C	50	22.4	936	22	Oligonucleotide D1
6	50	22.4	936	22	Oligonucleotide D1
C	50	22.4	936	22	Oligonucleotide D1
8	50	22.4	936	22	Oligonucleotide D1
C	50	22.4	936	22	Oligonucleotide D1
9	50	22.4	936	22	Oligonucleotide D1
C	50	22.4	936	22	Oligonucleotide D1
10	50	22.4	936	22	Oligonucleotide D1
C	50	22.4	936	22	Oligonucleotide D1
11	50	22.4	936	22	Oligonucleotide D2

12	50	22.4	936	22	AAF58262	Oligonucleotide D2	
C	13	50	22.4	936	22	AAF58262	Oligonucleotide D2
14	50	22.4	938	22	AAF58255	Oligonucleotide D1	
C	15	50	22.4	938	22	AAF58255	Oligonucleotide D1
C	16	47.4	21.3	244	22	AAF58238	Oligonucleotide D1
C	17	44	19.7	1153	21	AA339525	Arabidopsis thalia
18	44	19.7	80240	20	AAV83940	NC-contig derived	
C	19	44	19.7	80595	20	AAV83939	NC-contig derived
C	20	40.8	18.3	1152	21	AA46330	Arabidopsis thalia
C	21	38.8	17.4	513445	22	AA161373	Soybean 318013 reg
22	38.2	17.1	1119	8	AA71222	Sequence of soybea	
23	37.4	16.8	32190	22	AA162927	Human genomic DNA	
24	37	16.6	470	22	AAH93301	Plasmodium falci	
C	25	37	16.6	50000	21	AAA64139	Nucleotide sequenc
26	36.8	16.5	426	22	AAH93297	Arabidopsis thalia	
C	27	36.8	16.5	15577	19	AAV35616	SHOX gene prelinin
C	28	36.8	16.5	32367	19	AAV35620	Human SHOX (short
C	29	36.6	16.4	893	22	AAH03844	Human cDNA clone (
C	30	36.6	16.4	1602	22	AAH15369	Human cDNA sequenc
31	36.4	16.3	2418	13	AAQ27886	P.falci parum GBP13	
32	36.4	16.3	32249	22	AA162932	Human genomic DNA	
33	36.2	16.2	412	22	AAH93300	Plasmodium falci	
C	34	35.8	16.1	916	22	AA05274	Arabidopsis thalia
C	35	35.8	16.1	1538	21	AAA81567	N. meningitidis pa
C	36	35.8	16.1	335913	22	AA161371	Soybean 240017 reg
C	37	35.8	16.1	335913	22	AA161372	Soybean 240017 reg
C	38	35	15.7	41100	22	AA510873	Human genomic DNA
C	39	35	15.7	101786	21	AAF22293	BAC containing rep
40	34.8	15.6	15418	21	AA63785	Nucleotide sequenc	
41	34.8	15.6	168575	22	AAH21613	Human hypocretin r	
C	42	34.8	15.6	168575	22	AAH21613	Human hypocretin r
C	43	34.6	15.5	6038	21	AA243720	Human FSH-beta DNA
C	44	34.6	15.5	7622	21	AA243718	Human FSH-beta DNA
45	34.4	15.4	2451	20	AAV82439	Vascular endotheli	

#### ALIGNMENTS

RESULT 1  
AAD06999 standard; DNA; 416 BP.  
ID AAD06999 standard; DNA; 416 BP.  
AC AAD06999;  
DT 06-AUG-2001 (first entry)  
XX  
DE Left (3') border flanking region of elite event MS-B2.  
KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; ds.  
XX  
OS Chimeric - Agrobacterium sp.  
OS Chimeric - Brassica sp.  
FH Key Location/Qualifiers  
FT misc\_feature 1..193  
FT /tag= a  
FT /note= "Corresponds to T-DNA"  
FT misc\_feature 194..416  
FT /tag= b  
FT /note= "Corresponds to plant DNA"  
XX  
XX WO200131042-A2.  
XX  
XX 03-MAY-2001.  
XX  
XX 26-OCT-2000; 2000WO-EP10680.  
XX  
XX 29-OCT-1999; 99US-0430497.  
XX  
XX (AVET ) AVENTIS CROPS SCIENCE NV.

```

PI Weston B, De Beuckeleer M;
XX WPI; 2001-300517/31.
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX harboring specific transformation events, particularly by presence of
XX male-sterility gene, at specific location in its genome -
XX Claim 11; Page 52; 53pp; English.
XX The present invention relates to a transgenic Brassica plant or its
XX seed, cells or tissues, characterised by harbouring a specific
XX transformation event, particularly by the presence of a male-sterility
XX gene, at a specific location in the Brassica genome. Transgenic
XX Brassica plant is useful for producing a hybrid seed by crossing the
XX transgenic plant with a male-fertile Brassica plant and harvesting the
XX hybrid seed from the transgenic Brassica plant.
XX The present sequence is left (3') border flanking region of elite event
XX MS-B2.
XX Sequence 416 BP; 137 A; 72 C; 54 G; 152 T; 1 other;

Query Match 100.0%; Score 223; DB 22; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.4e-49;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaccataattataattataattataataactgaacacatggtgccctgctgttta 60
DB 194 ctaccataattataattataattataataactgaacacatggtgccctgctgttta 253
QY 61 catggatttcctcgctactattgtatacgtgtatataaccgtataatgtacatatatt 120
DB 254 catggatttcctcgctactattgtatacgtgtatataaccgtataatgtacatatatt 313
QY 121 tatatgaacatgattatgctgtgagttgttctcctcgtgaagagtttcaatatgtaatt 180
DB 314 tatatgaacatgattatgctgtgagttgttctcctcgtgaagagtttcaatatgtaatt 223
QY 181 ggtgaagagtcgaacaccccaaaatcatgaacaccccaaaactcgat 223
DB 374 ggtgaagagtcgaacaccccaaaatcatgaacaccccaaaactcgat 416

RESULT 2
AAD06997/c
ID AAD06997 standard; DNA; 415 BP.
AC AAD06997;
XX
XX 06-AUG-2001 (first entry)
XX
XX Right (5') border flanking region of elite event MS-B2.
XX
XX MS-B2 elite event; transgenic Brassica plant; transformation event;
XX male-sterility gene; ds.
XX
XX Chimeric - Agrobacterium sp.
XX Chimeric - Brassica sp.
XX
XX Key Location/Qualifiers
XX misc_feature 1..234
XX /*tag= a
XX /*note= "Corresponds to plant DNA"
XX misc_feature 235..415
XX /*tag= b
XX /*note= "Corresponds to T-DNA"
XX
XX WO200131042-A2.
XX
XX 03-MAY-2001.
XX
XX 26-OCT-2000; 2000WO-EP10680.

```

```

XX 29-OCT-1999; 99US-0430497.
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX Weston B, De Beuckeleer M;
XX WPI; 2001-300517/31.
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX harboring specific transformation events, particularly by presence of
XX male-sterility gene, at specific location in its genome -
XX Claim 11; Page 51; 53pp; English.
XX The present invention relates to a transgenic Brassica plant or its
XX seed, cells or tissues, characterised by harbouring a specific
XX transformation event, particularly by the presence of a male-sterility
XX gene, at a specific location in the Brassica genome. Transgenic
XX Brassica plant is useful for producing a hybrid seed by crossing the
XX transgenic plant with a male-fertile Brassica plant and harvesting the
XX hybrid seed from the transgenic Brassica plant.
XX The present sequence is right (5') border flanking region of elite event
XX MS-B2.
XX Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 85.1%; Score 189.8; DB 22; Length 415;
Best Local Similarity 98.2%; Pred. No. 1.1e-40;
Matches 213; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 ctaccataattataattataattataataac-tgaacacatggtgccctgctgttt 59
DB 222 ctaccataattataattataattataataacttgcaaacatggtgccctgctgttt 163
QY 60 acatgatttcctcgctactattgtatacgtgtatataaccgtataatgtacatatatt 119
DB 162 acatgatttcctcgctactattgtatacgtgtatataaccgtataatgtacatatatt 104
QY 120 ttatatgaacatgattatgctgtgagttgttctcctcgtgaagagtttcaatatgtaa 179
DB 103 ttatatgaacatgattatgctgtgagttgttctcctcgtgaagagtttcaatatgtaa 44
QY 180 tgggaagagtcgaacaccccaaaatcatgaacaccccaaa 216
DB 43 tgggaagagtcgaacaccccaaaatcatgaacaccccaaa 7

RESULT 3
AAF58238
ID AAF58238 standard; DNA; 244 BP.
XX
XX AAF58238;
XX
XX 24-APR-2001 (first entry)
XX
XX Oligonucleotide D1250.D1102.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX Synthetic.
XX
XX WO200107665-A2.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-US20476.
XX
XX 26-JUL-1999; 99US-0145695.
XX
XX 17-MAR-2000; 2000US-0190259.
XX

```

PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	Umek RM;
PI	WPI; 2001-159728/16.
XX	
DR	Nucleic acids containing electron-transfer group, useful as labels in
XX	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface -
PT	
XX	Example 4; Page 120; 159pp; English.
PS	The present invention relates to a composition comprising two nucleic
XX	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
PS	Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;
XX	
SQ	
	Query Match 22.6%; Score 50.4; DB 22; Length 244;
	Best Local Similarity 6.0%; Pred. No. 0.0002;
	Matches 12; Conservative 126; Mismatches 62; Indels 0; Gaps 0;
QY	6 ataattataattataataactgaaccatgtgccccctgcgtgtttcatcagg 65
	: :
DB	21 www 80
	: :
QY	66 atttcctcgctactatttgtagtgcgttatataccgtaataatgatcacatatatttat 125
	: :
DB	81 www 140
	: :
QY	126 gaacatgattaatgcttgagttgttctcatccgtaagagtccaatgaaatggtga 185
	: :
DB	141 www 200
	: :
QY	186 agagtcacaaaaccccaaatca 205
	: :
DB	201 wtgttaaagacacctca 220
	: :
RESULT 4	
AAF58252	
ID	AAF58252 standard; DNA; 936 BP.
AC	AAF58252;
XX	
DT	24-APR-2001 (first entry)
XX	Oligonucleotide D1835.
DE	
KW	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	Synthetic.
OS	
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PI	Umek RM;
XX	WPI; 2001-159728/16.
DR	
XX	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface -
XX	
PS	Example 6; Page 127; 159pp; English.
XX	The present invention relates to a composition comprising two nucleic
CC	acids each containing an electron-transfer group (ETM) having
CC	different redox potentials. The invention is used for electronic
CC	detection of nucleic acids, especially of substitutions (mismatches)
CC	and single-nucleotide polymorphisms, e.g. for genotyping,
CC	monitoring gene expression.
XX	
PS	Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
XX	
SQ	
	Query Match 22.4%; Score 50; DB 22; Length 936;
	Best Local Similarity 1.0%; Pred. No. 0.0003;
	Matches 2; Conservative 141; Mismatches 61; Indels 0; Gaps 0;
QY	6 ataattataattataataactgaaccatgtgccccctgcgtgtttcatcagg 65
	: :
DB	15 www 74
	: : : : ~~~~~~ :
QY	66 atttcctcgctactatttgtagtgcgttatataccgtaataatgatcacatatatttat 125
	: :
DB	75 www 134
	: :
QY	126 gaacatgattaatgcttgagttgttctcatccgtaagagtccaatgaaatggtga 185
	: :
DB	135 www 194
	: : : : ~~~~~~ :
QY	186 agagtcacaaaaccccaaatcatgaa 209
	: : : : ~~~~~~ :
DB	195 wgww 218
	: : : : ~~~~~~ :
RESULT 5	
AAF58252/c	
ID	AAF58252 standard; DNA; 936 BP.
XX	
AC	AAF58252;
XX	
DT	24-APR-2001 (first entry)
DE	Oligonucleotide D1835.
XX	
KW	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	Synthetic.
OS	
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PI	Umek RM;
XX	WPI; 2001-159728/16.
DR	
XX	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface -
XX	
PS	Example 6; Page 127; 159pp; English.
XX	



```
Matches 2; Conservative 141; Mismatches 61; Indels 0; Gaps 0;
Qy 6 ataattataataataataataactgaacacatggtgccccctgctttacatgg 65
Db 739 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
Qy 66 attctcgcgtactattgtgtacatgtgtatataacacgtataatgtacatatatttat 125
Db 679 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
Qy 126 gaacatgattaactgtgtgagtggttctcatcgcgtgaagagtttcaatatgtaatggtga 185
Db 619 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
Qy 186 agagtcacaaaccccaaatcatgaa 209
Db 559 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
RESULT 8
AAF58257
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
WPI; 2001-159728/16.
XX
Nucleic acids containing electron-transfer group, useful as labels in
hybridization assays, e.g. for genotyping, allowing repeat analyses on
a single surface
XX
Example 6; Page 127; 159pp; English.
XX
The present invention relates to a composition comprising two nucleic
acids each containing an electron-transfer group (ETM) having
different redox potentials. The invention is used for electronic
detection of nucleic acids, especially of substitutions (mismatches)
and single-nucleotide polymorphisms, e.g. for genotyping,
monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;
Query Match 22.4%; Score 50; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. NO. 0.00033;
Matches 2; Conservative 141; Mismatches 61; Indels 0; Gaps 0;
Qy 6 ataattataataataataataactgaacacatggtgccccctgctttacatgg 65
Db 15 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
```

```
Qy 66 attctcgcgtactattgtgtacatgtgtatataacacgtataatgtacatatatttat 125
Db 75 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
Qy 126 gaacatgattaactgtgtgagtggttctcatcgcgtgaagagtttcaatatgtaatggtga 185
Db 135 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
Qy 186 agagtcacaaaccccaaatcatgaa 209
Db 195 wgwwww wwwwww wwwwww wwwwww wwwwww wwwwww wwwwww wwwwww
RESULT 9
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
WPI; 2001-159728/16.
XX
Nucleic acids containing electron-transfer group, useful as labels in
hybridization assays, e.g. for genotyping, allowing repeat analyses on
a single surface
XX
Example 6; Page 127; 159pp; English.
XX
The present invention relates to a composition comprising two nucleic
acids each containing an electron-transfer group (ETM) having
different redox potentials. The invention is used for electronic
detection of nucleic acids, especially of substitutions (mismatches)
and single-nucleotide polymorphisms, e.g. for genotyping,
monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;
Query Match 22.4%; Score 50; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. NO. 0.00033;
Matches 2; Conservative 141; Mismatches 61; Indels 0; Gaps 0;
Qy 6 ataattataataataataataactgaacacatggtgccccctgctttacatgg 65
Db 739 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
Qy 66 attctcgcgtactattgtgtacatgtgtatataacacgtataatgtacatatatttat 125
Db 679 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW
Qy 126 gaacatgattaactgtgtgagtggttctcatcgcgtgaagagtttcaatatgtaatggtga 185
```







```

PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PS Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 22.4%; Score 50; DB 22; Length 938;
Best Local Similarity 1.0%; Pred. No. 0.00033;
Matches 2; Conservative 141; Mismatches 61; Indels 0; Gaps 0;

QY 6 ataattataattataactgaaccatggtgcccctgctgttacctgg 65
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 attctccgctactattgtatcggtatataaccgtataatgtacatatatttat 125
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 gaacatgattatgctgtgagttgtctcatccgcgaagtttcaatgtaattgga 185
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 agagtcacaaacccaaatcatgaa 209
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 559 agagtcacaaacccaaatcatgaa 536
Db : : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: February 15, 2002, 19:01:59
Job time: 20763 sec

```

---

```

PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PS Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 22.4%; Score 50; DB 22; Length 938;
Best Local Similarity 1.0%; Pred. No. 0.00033;
Matches 2; Conservative 141; Mismatches 61; Indels 0; Gaps 0;

QY 6 ataattataattataactgaaccatggtgcccctgctgttacctgg 65
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 15 www.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww 74
QY 66 attctccgctactattgtatcggtatataaccgtataatgtacatatatttat 125
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 www.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww 134
QY 126 gaacatgattatgctgtgagttgtctcatccgcgaagtttcaatgtaattgga 185
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 www.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww 194
QY 186 agagtcacaaacccaaatcatgaa 209
Db : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 195 wgwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww.wwww 218
Db : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
AAF58255/C
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:47:06 ; Search time 2553.1 Seconds  
(without alignments)  
2681.575 Million cell updates/sec

Title: US-09-698-903b-8

Perfect score: 415

Sequence: 1 gtcgagttgtgttcata.....cagctgtacattgcgtag 415

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_em.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Query Match	Length	DB ID	Description
1	415	100.0	415	6	AX127755
C 2	380.8	91.8	416	6	AX127757 Sequence
C 3	191.4	46.1	5865	6	AX127748 Sequence
C 4	191.4	46.1	7599	6	AX063413 Sequence
C 5	188.8	45.5	6539	6	E31991 Mutated bar
C 6	188.8	45.5	5548	6	A60109 Sequence 2
C 7	188.8	45.5	6548	6	A76916 Sequence 2
C 8	188.8	45.5	6548	6	AR098308 Sequence
C 9	188.8	45.5	6548	6	E31990 Mutated bar
C 10	188.8	45.5	7811	6	AR078675 Sequence
C 11	188.4	45.4	3200	6	I44104 Sequence 23
C 12	188.4	45.4	4832	6	AX172441 Sequence
C 13	188.4	45.4	4946	6	A60108 Sequence 1
C 14	188.4	45.4	4946	6	A76915 Sequence 1
C 15	188.4	45.4	4946	6	AR098307 Sequence
C 16	188.4	45.4	4946	6	AX172440 Sequence
C 17	188.4	45.4	5349	6	A71437 Sequence 7
C 18	188.4	45.4	5560	6	A60112 Sequence 5
C 19	188.4	45.4	5560	6	AR098311 Sequence
C 20	188.4	45.4	5865	6	AX127748 Sequence
C 21	188.4	45.4	7566	6	A24783 Sequence
C 22	188.4	45.4	7566	6	AR074388 Sequence
C 23	188.4	45.4	7639	6	A24782 Sequence
C 24	188.4	45.4	7639	6	AR074387 Sequence
C 25	182.4	44.0	1037	6	A10942 Nucleotide
C 26	182.4	44.0	1085	6	A10939 Nucleotide
C 27	182.4	44.0	1160	6	A10943 Nucleotide
C 28	179.4	43.2	1077	6	AX172463 Sequence
C 29	177.4	42.7	3201	6	I44103 Sequence 22
C 30	155	37.3	2476	12	TBI251013 Transform
C 31	155	37.3	3236	12	TBI251014 Transform
C 32	153	36.9	1186	6	A18051 DNA used as
C 33	153	36.9	1186	6	AR095107 Sequence
C 34	153	36.9	1186	6	AR098313 Sequence
C 35	153	36.9	1186	6	AX012338 Sequence
C 36	153	36.9	1186	6	I49886 Sequence 2
C 37	153	36.9	1186	6	I82374 Sequence 2
C 38	153	36.9	12095	12	BINHYGDNA
C 39	146	35.2	831	1	ATTNR7
C 40	146	35.2	878	1	ATTDNA
C 41	146	35.2	24595	1	ATACH5
C 42	146	35.2	24595	6	E00404
C 43	146	35.2	24595	6	E00546
C 44	146	35.2	194140	1	AF242881
C 45	116	28.0	249	12	ARGMTUB

#### ALIGNMENTS

#### RESULT 1

AX127755	AX127755	415 bp	DNA	PAT	15-MAY-2001
LOCUS	Sequence 8 from Patent WO0131042.				
DEFINITION	AX127755				
ACCESSION	AX127755.1	GI:14134402			
VERSION					
KEYWORDS	synthetic construct.				
SOURCE	synthetic construct.				
ORGANISM	artificial sequence.				
REFERENCE	1 (bases 1 to 415)				
AUTHORS	Weston, B. and de Beuckeleer, M.				
TITLE	Male-sterile brassica plants and methods for producing same				
JOURNAL	Patent: WO 0131042-A 8 03-MAY-2001;				
FEATURES	Aventis CropScience N.V. (BE)				
source	Location/Qualifiers				
	1..415				
	/organism="synthetic construct"				
	/db_xref="taxon:32630"				
	/note="5' border flanking region of elite event MS-B2"				

```
misc_feature 1..234
/Note="plant DNA"
misc_feature 235..415
/Note="T-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 415; DB 6; Length 415;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcgagtttggttcacgatttgggttttgactcttcaccattacattgaaactct 60
Db 1 gtcgagtttggttcacgatttgggttttgactcttcaccattacattgaaactct 60

QY 61 tacgagtagaacaactcacaagcatttaactgttcataataataatgtacattat 120
Db 61 tacgagtagaacaactcacaagcatttaactgttcataataataatgtacattat 120

QY 121 gtatatatacagtagatgagcagcatttaactgttcataataataatgtacattat 180
Db 121 gtatatatacagtagatgagcagcatttaactgttcataataataatgtacattat 180

QY 181 atggtttcagattatataataataataataataataataataataataataata 240
Db 181 atggtttcagattatataataataataataataataataataataataataata 240

QY 241 gaaaggcgaatttgatgttaattcccatcttgaaagaaatccatgtaaaagcagggggcacc 300
Db 241 gaaaggcgaatttgatgttaattcccatcttgaaagaaatccatgtaaaagcagggggcacc 300

QY 301 tgataaaatacaagtcaggtattatagtcgaagcaaaacataataatttattgacgag 360
Db 301 tgataaaatacaagtcaggtattatagtcgaagcaaaacataataatttattgacgag 360

QY 361 tttaattcagaataattttcaataactgtattatatacagtcggtacattcccgtag 415
Db 361 tttaattcagaataattttcaataactgtattatatacagtcggtacattcccgtag 415

RESULT 2
AX127757/c
LOCUS AX127757 416 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 10 from Patent WO0131042.
ACCESSION AX127757
VERSION AX127757.1 GI:14134404
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 416)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
Location/Qualifiers
source
1..416
/organism="synthetic construct"
/db_xref="taxon:32630"
/Note="3' border flanking region of elite event MS-B2"

misc_feature 1..193
/Note="T-DNA"
misc_feature 194..416
/Note="plant DNA"
BASE COUNT 137 a 72 c 54 g 152 t 1 others
ORIGIN

Query Match
Best Local Similarity 91.8%; Score 380.8; DB 6; Length 416;
Matches 404; Conservative 1; Mismatches 3; Indels 2; Gaps 2;

misc_feature 1..234
/Note="plant DNA"
misc_feature 235..415
/Note="T-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 415; DB 6; Length 415;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcgagtttggttcacgatttgggttttgactcttcaccattacattgaaactct 60
Db 1 gtcgagtttggttcacgatttgggttttgactcttcaccattacattgaaactct 60

QY 61 tacgagtagaacaactcacaagcatttaactgttcataataataatgtacattat 120
Db 61 tacgagtagaacaactcacaagcatttaactgttcataataataatgtacattat 120

QY 121 gtatatatacagtagatgagcagcatttaactgttcataataataatgtacattat 180
Db 121 gtatatatacagtagatgagcagcatttaactgttcataataataatgtacattat 180

QY 181 atggtttcagattatataataataataataataataataataataataataata 240
Db 181 atggtttcagattatataataataataataataataataataataataataata 240

QY 241 gaaaggcgaatttgatgttaattcccatcttgaaagaaatccatgtaaaagcagggggcacc 300
Db 241 gaaaggcgaatttgatgttaattcccatcttgaaagaaatccatgtaaaagcagggggcacc 300

QY 301 tgataaaatacaagtcaggtattatagtcgaagcaaaacataataatttattgacgag 360
Db 301 tgataaaatacaagtcaggtattatagtcgaagcaaaacataataatttattgacgag 360

QY 361 tttaattcagaataattttcaataactgtattatatacagtcggtacattcccgtag 415
Db 361 tttaattcagaataattttcaataactgtattatatacagtcggtacattcccgtag 415

RESULT 3
AX127748/c
LOCUS AX127748 5865 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 1 from Patent WO0131042.
ACCESSION AX127748
VERSION AX127748.1 GI:14134395
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 5865)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 1 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
Location/Qualifiers
source
1..5865
/organism="synthetic construct"
/db_xref="taxon:32630"
/Note="T-DNA of plasmid pCO113"
BASE COUNT 1849 a 1095 c 1149 g 1772 t
ORIGIN

Query Match 46.1%; Score 191.4; DB 6; Length 5865;
Best Local Similarity 99.5%; Pred. No. 5.7e-25;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 223 gatgtacatggccgataaaggaagcaattttagatgttaattcccatcttgaagaaa 282
Db 5813 gatgtacatggccgataaaggaagcaattttagatgttaattcccatcttgaagaaa 5754

QY 283 tatagtttaataattttattgataaaatacaagtcaggtattatagtcgaagcaaa 342
Db 5753 TATAGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACA 5694

QY 343 taaattattgacgaagtttaaatcagaataatttcaataactgattatcagctgg 402
Db 5693 TAAATTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATACGCTGG 5634

QY 403 tacattgcccgtag 415
Db 5633 TACATTGCCGTAG 5621
```

```
RESULT 4
AX063413/c
LOCUS AX063413 7599 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 5 from Patent WO0100833.
ACCESSION AX063413
VERSION AX063413.1 GI:12541201
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 7599)
AUTHORS Hoffmann,B., Mollier,P. and Pelletier,G.
TITLE Promoter expressed specifically in the cells of plant roots,
recombinant vectors and host cells comprising same and transgenic
plants obtained
JOURNAL Patent: WO 0100833-A 5 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES
source Location/Qualifiers
1..7599
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="ADN-T de PK85"
BASE COUNT 1972 a 1938 c 1937 g 1752 t
ORIGIN
Query Match 46.1%; Score 191.4; DB 6; Length 7599;
Best Local Similarity 99.5%; Pred. No. 5.6e-25;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 223 gatgtacatggccgataagaagcaattttagatgttaattcccatcttgaagaaa 282
|||||
Db 6771 GATGTACATGTCGATAGAAAGCAATTGTAGATGTTAATCCCATCTTGAAGAAA 6712
Qy 283 tatagttcaaatattattgtataataaacaagtcaggattattatagtcacaagcaaaaaca 342
|||||
Db 6711 TATAGTTTAAATATTATTGTGATAAATAACAAGTCAGGTATTATATAGTCCAAGCAAAAAACA 6652
Qy 343 taaattattatgcgaatttaaaatcagaataatttcaataaactgattatcagctgg 402
|||||
Db 6651 TAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATTAACATGATTATATACAGCTGG 6592
Qy 403 tacattgccgtag 415
|||||
Db 6591 TACATTGCCGTAG 6579
RESULT 5
E31991/c
LOCUS E31991 6539 bp DNA PAT 07-FEB-2001
DEFINITION Mutated barnase gene and transgenic plant thereof.
ACCESSION E31991
VERSION E31991.1 GI:13021588
KEYWORDS JP 2000041682-A/4.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 6539)
AUTHORS Kazuyuki,H.F.N.N.
TITLE Mutated barnase gene and transgenic plant thereof
JOURNAL Patent: JP 2000041682-A 4 15-FEB-2000;
JAPAN TOBACCO INC
COMMENT OS Escherichia coli LE392
PN JP 2000041682-A/4
PD 15-FEB-2000
PF 04-AUG-1998 JP 1998220060
PR KAZUYUKI HAMADA,FUMIO NAKAKIDO
PI C12N15/09,A01H5/00,C12N5/10,C12N9/22/(C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00,(C12N5/00,C12R1:91)
CC
```

```

FH Key Location/Qualifiers
FT source 1..6539
/organism="Escherichia coli LE392".
FEATURES
source Location/Qualifiers
1..6539
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1755 a 1578 c 1519 g 1687 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6539;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 224 atgtacatggccgataagaagcaattttagatgttaattcccatcttgaagaaaat 283
|||||
Db 6533 ACGTACATGGTCGATAGAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAAAAT 6474
Qy 284 atagtttaaatatttattgataaataacaagtcaggattattatagtcacaagcaaaaacat 343
|||||
Db 6473 ATAGTTTAAATATTATTGATAAATNACAAGTCAGGTATTATAGTCCAAGCAAAAACAT 6414
Qy 344 aaattattatgcgaagttaaaatcagaataatttcaataaactgattatcagctggt 403
|||||
Db 6413 AAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 6354
Qy 404 acattgccgtag 415
|||||
Db 6353 ACATTGCCGTAG 6342
RESULT 6
A60109/c
LOCUS A60109 6548 bp DNA circular PAT 22-OCT-1999
DEFINITION Sequence 2 from Patent WO9706267.
ACCESSION A60109
VERSION A60109.1 GI:3715125
KEYWORDS
SOURCE Plasmid PTS172.
ORGANISM Plasmids.
REFERENCE 1 (bases 1 to 6548)
AUTHORS De,B.M.
TITLE Genetic transformation using a PARP inhibitor
JOURNAL Patent: WO 9706267-A 2 20-FEB-1997;
PLANT GENETIC SYSTEMS NV (BE)
FEATURES
source Location/Qualifiers
1..6548
/organism="Plasmid PTS172"
/db_xref="taxon:106340"
BASE COUNT 1756 a 1579 c 1523 g 1690 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 224 atgtacatggccgataagaagcaattttagatgttaattcccatcttgaagaaaat 283
|||||
Db 6542 ACGTACATGGTCGATAGAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAAAAT 6483
Qy 284 atagtttaaatatttattgataaataacaagtcaggattattatagtcacaagcaaaaacat 343
|||||
Db 6482 ATAGTTTAAATATTATTGATAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACAT 6423
Qy 344 aaattattatgcgaagttaaaatcagaataatttcaataaactgattatcagctggt 403
|||||
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 6363
Qy 404 acattgccgtag 415
|||||
```

QY 224 atgtacatggccgataagaaaaaggcaatttgtagatgttaattcccatcttgaagaaaat 283



```
PLANT GENETIC SYSTEMS NV (BE)
FEATURES             source
  LOCUS              1..4946
  DEFINITION         /organism="Transformation vector pTHW107"
  ACCESSION          /db_xref="taxon:126810"
  VERSION            1569 a 891 c 963 g 1523 t
  KEYWORDS            BASE COUNT 1569 a 891 c 963 g 1523 t
  ORIGIN
Query Match          45.4%; Score 188.4; DB 6; Length 4946;
Best Local Similarity 99.5%; Pred. No. 2e-24;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 226 gtacatggccgataagaagaaggcaatttggatgttaattcccatcttgaagaataat 285
DB 54 gtacatggcgtcgataagaagaaggcaatttggatgttaattcccatcttgaagaataat 113
QY 286 agtttaaatattattgtataaataacaagtcagggtattatagtcgaagcaaaacataa 345
DB 114 agtttaaatattattgtataaataacaagtcagggtattatagtcgaagcaaaacataa 173
QY 346 atttattgatgcaagtttaaatattcaataaattcaataaactgattatcagctggtac 405
DB 174 atttattgatgcaagtttaaatattcaataaattcaataaactgattatcagctggtac 233
QY 406 attgccgtag 415
DB 234 attgccgtag 243
RESULT 14
LOCUS              A76915
DEFINITION         Sequence 1 from Patent EP0757102.
ACCESSION          A76915
VERSION            A76915.1 GI:6088712
KEYWORDS            Transformation vector pTHW107.
  ORIGIN            Transformation vector pTHW107.
  REFERENCE         1 (bases 1 to 4946)
  AUTHORS            De,B.M.
  TITLE             GENETIC TRANSFORMATION USING A PARP INHIBITOR
  JOURNAL            Patent: EP 0757102-A 1 05-FEB-1997;
  FEATURES            PLANT GENETIC SYSTEMS NV (BE)
    source            Location/Qualifiers
      1..4946
      /organism="Transformation vector pTHW107"
      /db_xref="taxon:126810"
BASE COUNT          1569 a 891 c 963 g 1523 t
ORIGIN
Query Match          45.4%; Score 188.4; DB 6; Length 4946;
Best Local Similarity 99.5%; Pred. No. 2e-24;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 226 gtacatggccgataagaagaaggcaatttggatgttaattcccatcttgaagaataat 285
DB 54 gtacatggcgtcgataagaagaaggcaatttggatgttaattcccatcttgaagaataat 113
QY 286 agtttaaatattattgtataaataacaagtcagggtattatagtcgaagcaaaacataa 345
DB 114 agtttaaatattattgtataaataacaagtcagggtattatagtcgaagcaaaacataa 173
QY 346 atttattgatgcaagtttaaatattcaataaattcaataaactgattatcagctggtac 405
DB 174 atttattgatgcaagtttaaatattcaataaattcaataaactgattatcagctggtac 233
QY 406 attgccgtag 415
DB 234 attgccgtag 243
```

Search completed: February 15, 2002, 18:47:39  
Job time: 20098 sec





**THIS PAGE BLANK (USPTO)**



BASE COUNT 7 a 6 c 5 g 3 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||

Db 1 GTAACATAGATGACACCGCGC 21

## RESULT 2

AR110601/c AR110601 249 bp DNA PAT 14-FEB-2001

LOCUS Sequence 8 from patent US 6114608.

ACCESSION AR110601

VERSION AR110601.1 GI:12826877

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 249)

AUTHORS Mettler, I.J., Dietrich, P.S. and Sinibaldi, R.M.

TITLE Nucleic acid construct comprising bacillus thuringiensis cryIaB

JOURNAL gene

FEATURES Patent: US 6114608-A 8 05-SEP-2000;

Location/Qualifiers

1..249

/organism="unknown"

BASE COUNT 79 a 38 c 47 g 85 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 3

AR151000/c AR151000 249 bp DNA PAT 08-AUG-2001

LOCUS Sequence 8 from patent US 6229075.

ACCESSION AR151000

VERSION AR151000.1 GI:15115591

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 249)

AUTHORS Mettler, I.J., Plaisted, D.C., Grier, S.L., Houghton, W. and

Gardiner, M.

TITLE Inbred maize line R412H

JOURNAL Patent: US 6229075-A 8 08-MAY-2001;

Location/Qualifiers

1..249

/organism="unknown"

BASE COUNT 79 a 38 c 47 g 85 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 4

AR152423/c AR152423 249 bp DNA PAT 08-AUG-2001

LOCUS Sequence 8 from patent US 6232533.

ACCESSION AR152423

VERSION AR152423.1 GI:15118473

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 249)

AUTHORS Mettler, I.J., Plaisted, D.C., Grier, S.L., Houghton, W. and

Gardiner, M.

TITLE Inbred maize line R372H

JOURNAL Patent: US 6232533-A 8 15-MAY-2001;

Location/Qualifiers

1..249

/organism="unknown"

BASE COUNT 79 a 38 c 47 g 85 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 5

AR152432/c AR152432 249 bp DNA PAT 08-AUG-2001

LOCUS Sequence 8 from patent US 6232534.

ACCESSION AR152432

VERSION AR152432.1 GI:15118482

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 249)

AUTHORS Mettler, I.J., Plaisted, D.C., Grier, S.L., Houghton, W. and

Gardiner, M.

TITLE Inbred maize line R660H

JOURNAL Patent: US 6232534-A 8 15-MAY-2001;

Location/Qualifiers

1..249

/organism="unknown"

BASE COUNT 79 a 38 c 47 g 85 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 6

AR110597/c AR110597 261 bp DNA PAT 14-FEB-2001

LOCUS Sequence 4 from patent US 6114608.

ACCESSION AR110597

VERSION AR110597.1 GI:12826873

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

```
Unclassified.
REFERENCE 1 (bases 1 to 261)
AUTHORS Mettler,I.J., Dietrich,P.S and Sinibaldi,R.M.
TITLE Nucleic acid construct comprising bacillus thuringiensis crylab
gene
JOURNAL Patent: US 6114608-A 4 05-SEP-2000;
FEATURES Location/Qualifiers
source 1..261
BASE COUNT 83 a 41 c 50 g 87 t
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21
|||||
Db 253 GTAACATAGATGACACGC GC 233

RESULT 9
AR152428/c
LOCUS AR152428 261 bp DNA
DEFINITION Sequence 4 from patent US 6232534.
ACCESSION AR152428
VERSION AR152428.1 GI:15118478
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 261)
AUTHORS Mettler,I.J., Plaisted,D.C., Grier,S.L., Houghton,W. and Gardiner,M.
TITLE Inbred maize line R660H
JOURNAL Patent: US 6232534-A 4 15-MAY-2001;
FEATURES Location/Qualifiers
source 1..261
BASE COUNT 83 a 41 c 50 g 87 t
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21
|||||
Db 253 GTAACATAGATGACACGC GC 233

RESULT 10
E01312/c
LOCUS E01312 280 bp DNA
DEFINITION 3' Untranslated region of nopalin synthetase gene.
ACCESSION E01312
VERSION E01312.1 GI:2169571
KEYWORDS JP 1987201527-A/4.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 280)
AUTHORS Rojiyaa,E.B., Robaato,T.F. and Suteilibun,J.R.
TITLE PROTECTION OF PLANT FROM VIRUS INFECTION
JOURNAL Patent: JP 1987201527-A 4 05-SEP-1987;
COMMENT MONSANTO CO, WASHINGTON UNIV
OS Agrobacterium
PN JP 1987201527-A/4
PD 05-SEP-1987
PF 29-OCT-1986 JP 1986258063
PR 29-OCT-1985 US 85 792389, 27-MAR-1986 US 86 844918, PR
09-OCT-1986 US 86 917027
PI ROJIYAA ENU BIICHII, ROBAATO TEII FURAREI,
PI SUTEILIBUN JII ROJIYAASU
PC A01H5/00,A01H1/00,C12N1/20,C12N15/00//A01N65/00,C07K13/00, PC
(c12n1/20,
PC C12R1:01);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: NO;
FH Key Location/Qualifiers
```

```

EH
FT
FEATURES
  source
    polyA_signal >1..<281.
    Location/Qualifiers
      1..280
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT      83 a 49 c 58 g 90 t
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 280;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtaacatagatgacacgcgc 21
    |||||
Db 246 GTAACATAGATGACACGC GC 226

RESULT 11
RICL429B/C
LOCUS      288 bp DNA
DEFINITION Rice genomic DNA, L429B, sequence tagged site.
ACCESSION D25449
VERSION D25449.1 GI:436695
KEYWORDS STS; Not a linking clone; RFLP; Rice.
SOURCE Oryza sativa (strain:Nipponbare) DNA.
ORGANISM Oryza sativa
          Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 288)
AUTHORS Minobe,Y.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1993) to the DDBJ/EMBL/GenBank databases. Yuzo
Minobe, National Institute of Agrobiological Resources, Rice Genome
Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305, Japan
(E-mail:MINOBE@res.riken.go.jp, Tel:0298-38-7441,
Fax:0298-38-7468)
REFERENCE 2 (bases 1 to 288)
AUTHORS Minobe,Y.
TITLE Nucleotide sequence of sequence tagged site from japonica rice
Nipponbare as an RFLP marker
JOURNAL Unpublished (1993)
COMMENT Submitted (18-Nov-1993) to DDBJ by:
Yuzo Minobe
Dept.. Rice Genome Research Program
National Institute of Agrobiological Resources
Kannondai 2-1-2
Tsukuba, Ibaraki
Japan
Phone: 0298-38-7441
Fax: 0298-38-7468
PROJECT "RGP".
FEATURES
  source
    Location/Qualifiers
      1..288
        /organism="Oryza sativa"
        /strain="Nipponbare"
        /db_xref="taxon:4530"
BASE COUNT      77 a 61 c 80 g 68 t 2 others
ORIGIN
Query Match      100.0%; Score 21; DB 11; Length 288;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtaacatagatgacacgcgc 21
    |||||
Db 113 GTAACATAGATGACACGC GC 93

RESULT 12
ABA7623
LOCUS      318 bp DNA
DEFINITION Agrobacterium tumefaciens nos terminator with artificial insert of
40 bp.
ACCESSION AJ007623
VERSION AJ007623.1 GI:3319860
KEYWORDS nos gene; terminator.
SOURCE Agrobacterium tumefaciens.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE 1 (bases 1 to 318)
AUTHORS Hardegger,M.R., Brodmann,P. and Herrmann,A.
TITLE Quantitative detection of the 35S promoter and the NOS terminator
using quantitative competitive PCR
JOURNAL 2. Lebensm.-Unters. -Forsch., A Eur. Food Res. Technol. 209, 83-87
(1999)
REFERENCE 2 (bases 1 to 318)
AUTHORS Hardegger,M.R.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1998) Hardegger M.R., Lebensmittel, Kantionales
Laboratorium Basel-Stadt, Kannenfeldstrasse 2, CH/4012 Basel,
SWITZERLAND
FEATURES
  source
    Location/Qualifiers
      1..318
        /organism="Agrobacterium tumefaciens"
        /plasmid="Ti"
        /db_xref="taxon:358"
      1..318
        /gene="nos"
      1..318
        /gene="nos"
        /note="insert location 142-181"
BASE COUNT      102 a 64 c 55 g 97 t
ORIGIN
Query Match      100.0%; Score 21; DB 1; Length 318;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtaacatagatgacacgcgc 21
    |||||
Db 16 GTAACATAGATGACACGC GC 36

RESULT 13
ABA7624
LOCUS      358 bp DNA
DEFINITION Agrobacterium tumefaciens nos terminator with artificial insert of
80 bp.
ACCESSION AJ007624
VERSION AJ007624.1 GI:3319861
KEYWORDS nos gene; terminator.
SOURCE Agrobacterium tumefaciens.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE 1 (bases 1 to 358)
AUTHORS Hardegger,M.R., Brodmann,P. and Herrmann,A.
TITLE Quantitative detection of the 35S promoter and the NOS terminator
using quantitative competitive PCR
JOURNAL 2. Lebensm.-Unters. -Forsch., A Eur. Food Res. Technol. 209, 83-87
(1999)
REFERENCE 2 (bases 1 to 358)
AUTHORS Hardegger,M.R.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1998) Hardegger M.R., Lebensmittel, Kantionales
Laboratorium Basel-Stadt, Kannenfeldstrasse 2, CH/4012 Basel,
SWITZERLAND
FEATURES
  source
    Location/Qualifiers
      1..358
        /organism="Agrobacterium tumefaciens"

```

```
gene /plasmid="ti"
terminator /db_xref="taxon:358"
1. .358
/gene="nos"
1. .358
/gene="nos"
BASE COUNT 114 a 75 c 63 g 106 t
ORIGIN /note="Insert location 142-221; direct repeat"

Query Match 100.0%; Score 21; DB 1; Length 358;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 16 GTAACATAGATGACACGC GC 36

RESULT 14
A71436/c A71436 563 bp DNA PAT 07-MAY-1999
LOCUS Sequence 6 from Patent WO9810081.
DEFINITION A71436
ACCESSION A71436
VERSION A71436.1 GI:4775049
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 563)
AUTHORS Michiels, F. and Williams, M.
TITLE IMPROVED BARSTAR GENE
JOURNAL Patent: WO 9810081-A 6 12-MAR-1998;
MICHIELS FRANK (BE)
FEATURES
source
1. .563
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 152 a 133 c 145 g 133 t
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 563;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 538 GTAACATAGATGACACGC GC 518

RESULT 15
AX014764/c AX014764 838 bp DNA PAT 07-SEP-2000
LOCUS Sequence 19 from Patent WO9953053.
DEFINITION AX014764
ACCESSION AX014764
VERSION AX014764.1 GI:10041035
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 838)
AUTHORS Lamberty, M., Hofmann, J., Bulet, P. and Brookhart, G. L.
TITLE Gene coding for heliomycin and use thereof
JOURNAL Patent: WO 9953053-A 19 21-OCT-1999;
LAMBERTY MIREILLE (FR); HOFMANN JULES (FR); BULET PHILIPPE (FR);
RHONE-POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)
FEATURES
source
1. .838
/organism="synthetic construct"
/db_xref="taxon:32630"
promoter
```

```
misc_structure 533. .568
terminator 569. .832
BASE COUNT 296 a 134 c 175 g 233 t
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 838;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 822 GTAACATAGATGACACGC GC 802

Search completed: February 15, 2002, 18:46:25
Job time: 20024 sec
```

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 13:15:56 ; Search time 868.33 Seconds  
(without alignments)  
20.734 Million cell updates/sec

Title: US-09-698-903B-2

Perfect score: 21  
Sequence: 1 gtaacatagatgacaccgcgc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	22	PCR primer MDB355
C 2	21	100.0	249	21	AA075465
C 3	21	100.0	249	22	AA063113
C 4	21	100.0	249	22	AA06367
C 5	21	100.0	249	22	AAF89824
C 6	21	100.0	249	22	AAF89833
C 7	21	100.0	253	21	AA251175
C 8	21	100.0	254	21	AA298689
C 9	21	100.0	257	21	AA251653
C 10	21	100.0	261	21	AA075461
C 11	21	100.0	261	22	AA06309

C 12	21	100.0	261	22	AA06363	3' nopaline synthase
C 13	21	100.0	261	22	AAF89820	Nucleotide sequenc
C 14	21	100.0	261	22	AAF89829	Nucleotide sequenc
C 15	21	100.0	263	18	AA066536	Nopaline synthase
C 16	21	100.0	270	20	AA066467	NOS terminator seq
C 17	21	100.0	270	22	AA089659	Tomato spotted wll
C 18	21	100.0	563	19	AA023238	Part of plasmid pl
C 19	21	100.0	900	20	AA087748	Vector plasmid psi
C 20	21	100.0	1138	18	AA085665	PMJBI DNA sequence
C 21	21	100.0	1138	18	AA059713	Plasmid PMJBI used
C 22	21	100.0	1138	20	AA078852	Plasmid PMJBI Ecor
C 23	21	100.0	1559	19	AA064071	Arabidopsis thalia
C 24	21	100.0	1630	21	AA01014	Expression cassett
C 25	21	100.0	1690	20	AA060384	Nematode-responsiv
C 26	21	100.0	1829	13	AA029293	35S CamV promoter-
C 27	21	100.0	1863	13	AA021191	Sequence of chimera
C 28	21	100.0	1863	15	AA074261	Tomato-tobacco end
C 29	21	100.0	1923	14	AA057524	OBPIIT sequence.
C 30	21	100.0	1949	20	AA05602	Nucleotide sequenc
C 31	21	100.0	2115	19	AA044279	Oleasin-hirudin fu
C 32	21	100.0	2145	18	AA086752	Raspberry drul gen
C 33	21	100.0	2275	22	AA086440	Oligonucleotide #1
C 34	21	100.0	2319	19	AA064070	Arabidopsis thalia
C 35	21	100.0	2345	19	AA054016	Nucleotide sequenc
C 36	21	100.0	2345	20	AA084450	T-DNA sequence of
C 37	21	100.0	2345	22	AA068664	Agrobacterium tune
C 38	21	100.0	2378	21	AA01016	Expression cassett
C 39	21	100.0	2436	21	AA01018	Expression cassett
C 40	21	100.0	2543	22	AA057334	DNA construct codi
C 41	21	100.0	2709	19	AA044284	Oleasin-protein A
C 42	21	100.0	2917	18	AA086755	Raspberry drul gen
C 43	21	100.0	3069	22	AA011579	PND3018 repression
C 44	21	100.0	3121	22	AA011578	PND3008 activation
C 45	21	100.0	3183	17	AA042919	SAG12-1 promoter,

#### ALIGNMENTS

RESULT 1  
AAD06991  
ID AAD06991 standard; DNA; 21 BP.  
XX  
AC AAD06991;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE PCR primer MDB355 to amplify a fragment of pTCO113.  
XX  
KW Plasmid pTCO113; transgenic Brassica plant; transformation event;  
KW male-sterility gene; tobacco; PCR primer; probe; ss.  
XX  
OS Chimeric - Bacillus amyloliquefaciens.  
OS Chimeric - Nicotiana tabacum.  
XX  
PN WO200131042-A2.  
XX  
PD 03-MAY-2001.  
XX  
PF 26-OCT-2000; 2000WO-EPI0680.  
XX  
PR 29-OCT-1999; 99US-0430497.  
XX  
PA (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
PI Weston B, De Beuckeleer M;  
XX  
DR WPI; 2001-300517/31.  
XX  
PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -

PS Claim 1; Page 26; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is PCR primer MDB355 used to amplify a fragment of  
 CC plasmid pTCO113. The primer is also used as a probe. The amplified  
 CC fragment comprises coding region of barnase from  
 CC Bacillus amyloliquefaciens and promoter region of the anther-specific  
 CC gene TA29 from Nicotiana tabacum. This primer corresponds to position  
 CC 2667-2687 of plasmid pTCO113.

XX Sequence 21 BP; 7 A; 6 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.057;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
 Db 1 gtaacatagatgacacgcgc 21

RESULT 2

AAA75465/c  
 ID AAA75465 standard; DNA; 249 BP.

AC AAA75465;

XX 15-JAN-2001 (first entry)

XX Nucleotide sequence of a nopaline synthase terminator.

XX 35S promoter; alcohol dehydrogenase intron; nopaline synthase terminator;  
 KW insecticidal Cry IAB protein toxin; transgenic plant; fungal toxin;  
 KW aflatoxin; baculovirus; Lepidopteran insect; glufosinate; ds.

XX Unidentified.

XX US6114608-A.

XX 05-SEP-2000.

XX 13-MAR-1998; 98US-0042426.

XX 14-MAR-1997; 97US-0109808.

XX (NOVS ) NOVARTIS AG.

PI Dietrich PS, Mettler IJ, Sinibaldi RM;

XX WPI; 2000-586487/55.

XX Novel DNA construct comprising an expression cassette having a  
 PT functional constitutive promoter, operably linked to a maize alcohol  
 PT dehydrogenase intron, gene encoding Cry IAB protein and a terminator  
 XX Claim 2; Column 25-26; 24pp; English.

XX The present sequence represents a nopaline synthase terminator, which is  
 CC used to create the construct of the invention. The specification  
 CC describes a nucleic acid construct, comprising an expression cassette  
 CC containing a cauliflower mosaic virus (CaMV) 35S promoter, a maize  
 CC alcohol dehydrogenase intron, a DNA molecule encoding an insecticidal  
 CC Cry IAB protein toxin and a NOS (nopaline synthase) terminator in  
 CC operable linkage. The nucleic acid constructs are useful for  
 CC producing transgenic plants such as maize, wheat, barley, sorghum and  
 CC rice. Preferably, maize plants such as field corn, sweet corn, white

CC corn, silage corn or popcorn are produced. The transgenic maize is  
 CC used for preparing food materials with reduced levels of fungal toxins,  
 CC e.g. aflatoxins. The transgenic plants express a baculovirus crystal  
 CC protein toxin, which is effective against Lepidopteran insects at  
 CC relatively high levels and further provides resistance to the  
 CC non-selective herbicide glufosinate.

SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 249;  
 Best Local Similarity 100.0%; Pred. No. 0.085;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
 Db 247 GTAACATAGATGACACGC GC 227

RESULT 3

AAD06313/c  
 ID AAD06313 standard; DNA; 249 BP.

XX AAD06313;

XX 10-AUG-2001 (first entry)

XX Nopaline synthase terminator of Pat gene expression cassette.

XX Transgenic maize; expression cassette; 35S promoter; nopaline synthetase;  
 KW NOS terminator; alcohol dehydrogenase intron; CryIAB toxin;  
 KW foodstuffs preparation; phosphinothricin acetyl transferase; PAT; ds.

XX Unidentified.

XX US6222104-B1.

XX 24-APR-2001.

XX 13-APR-1999; 99US-0291238.

XX 09-NOV-1994; 94US-0336627.

XX 22-AUG-1996; 96US-0716836.

XX 14-MAR-1997; 97US-0818573.

XX 13-MAR-1998; 98US-0042426.

XX (NOVS ) NOVARTIS AG.

XX Mettler IJ, Krier M, Mies D;

XX WPI; 2001-327266/34.

XX Novel transgenic maize seed for hybrid maize plant production,  
 PT comprising expression cassette linked operably with CaMV 35S promoter,  
 PT alcohol dehydrogenase intron, insecticidal CryIAB toxin gene or NOS  
 PT terminator

XX Claim 3; Column 25-26; 24pp; English.

XX The present invention relates to transgenic maize seed for hybrid maize  
 CC plant production, comprising an expression cassette linked operably with  
 CC constitutive Cauliflower mosaic virus (CaMV) 35S promoter, an alcohol  
 CC dehydrogenase intron, a DNA encoding an insecticidal  
 CC Bacillus thuringiensis (Bt) CryIAB toxin and a nopaline synthetase (NOS)  
 CC terminator. The transgenic maize seed is used in maize cultivation  
 CC methods for propagating hybrid maize seeds and for growing hybrid maize  
 CC plants. The maize is also used in foodstuffs preparation for animal or  
 CC human consumption. The inbred maize line NP948 of the invention has  
 CC reduced level of fungal toxins, hence suitable for foodstuffs  
 CC preparation. Yield is high. Plant health is improved due to reduced  
 CC stalk rot. Grain test weight is high and the rate of grain dry down is  
 CC reduced. The present sequence is nopaline synthetase terminator of  
 CC phosphinothricin acetyl transferase (PAT) gene expression cassette.

XX SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;  
 Best Local Similarity 100.0%; Pred. No. 0.085;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
 |||||  
 Db 247 GTRACATAGATGACACGC GC 227

RESULT 4  
 AAD06367/c  
 ID AAD06367 standard; DNA; 249 BP.  
 AC AAD06367;  
 XX  
 DT 10-AUG-2001 (first entry)  
 DE  
 DE 3' nopaline synthase terminator of Pat gene expression cassette.  
 XX  
 XX Maize: expression cassette; 35S promoter; alcohol dehydrogenase intron;  
 KW crystal lab; Crylab toxin; phosphinothricin acetyl transferase; PAT;  
 KW nopaline synthase; NOS terminator; aflatoxin; food material;  
 KW sweet corn; human food; Adhl-1S; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN US6229075-B1.  
 PD 08-MAY-2001.  
 XX  
 XX 11-JUN-1999; 99US-0330760.  
 PF  
 XX 14-MAR-1997; 97US-0109808.  
 PR 13-MAR-1998; 98US-0042426.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;  
 XX WPI; 2001-342708/36.  
 DR  
 XX New maize (Zea mays) inbred line R412H (ATCC 209675), useful for  
 PT producing corn with reduced levels of toxins (e.g. the fungal  
 PT aflatoxin) and for preparing food materials for human or animal  
 PT consumption -  
 XX  
 XX Claim 3; Column 21-22; 30pp; English.  
 PS  
 XX The present invention relates to maize inbred line R412H which comprises  
 CC a nucleic acid construct with two cassettes, which are transcribed in the  
 CC same direction. The expression cassette comprises a Cauliflower mosaic  
 CC virus (CaMV) 35S constitutive promoter operably linked to a maize alcohol  
 CC dehydrogenase intron, a DNA sequence of a gene encoding a  
 CC Bacillus thuringiensis (Bt) crystal lab (CryIAb) toxin or  
 CC phosphinothricin acetyl transferase (PAT) and a nopaline synthase (NOS)  
 CC terminator functional in plants. The maize inbred line R412H is useful  
 CC for producing corn with reduced levels of toxins (e.g. the fungal  
 CC aflatoxin). This maize line is particularly useful for preparing food  
 CC materials for human or animal consumption, e.g. sweet corn for  
 CC packaging or fresh use as human food, or grain or silage made from field  
 CC corn. The present sequence is 3' NOS terminator of Pat gene expression  
 CC cassette, which is related to the invention.  
 XX  
 XX SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;  
 Best Local Similarity 100.0%; Pred. No. 0.085;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
 |||||  
 Db 247 GTRACATAGATGACACGC GC 227

RESULT 5  
 AAF89824/c  
 ID AAF89824 standard; DNA; 249 BP.  
 XX  
 AC AAF89824;  
 XX  
 DT 23-JUL-2001 (first entry)  
 DE  
 DE Nucleotide sequence of a 3' terminator from nopaline synthetase.  
 XX  
 XX Maize; inbred line R660H; 35S constitutive promoter; Crylab protein;  
 KW alcohol dehydrogenase; phosphinothricin acetyl transferase; chromosome 8;  
 KW Z1B3; UMC150a; Lepidoptera; glufoisinate; sweet corn; fungal toxin; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN US6232533-B1.  
 XX  
 PD 15-MAY-2001.  
 XX  
 PF 09-JUN-1999; 99US-0328473.  
 XX  
 XX 14-MAR-1997; 97US-0109808.  
 PR 13-MAR-1998; 98US-0042426.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;  
 XX WPI; 2001-335091/35.  
 DR  
 XX Novel seed of maize inbred line R372H, useful for producing inbred  
 PT maize plants which on crossing with other different maize plants  
 PT produce hybrid maize seeds and plants having resistance to insects and  
 PT herbicide -  
 XX  
 XX Claim 3; Column 19-21; 29pp; English.  
 PS  
 XX The specification describes seed and plants of maize inbred line R660H.  
 CC The seed comprises two cassettes. The first cassette comprises a  
 CC cauliflower mosaic virus (CaMV) 35S constitutive promoter operably linked  
 CC to a maize alcohol dehydrogenase intron, a DNA sequence of a gene  
 CC encoding a CryIAb protein, and a terminator functional in plants. The  
 CC second cassette comprises a CaMV 35S promoter which functions in plant  
 CC cells operably linked to a maize alcohol dehydrogenase intron, a DNA  
 CC sequence of a gene encoding for phosphinothricin acetyl transferase, and  
 CC a terminator functional in plant. The two cassettes are transcribed in  
 CC the same direction and the nucleic acid construct is incorporated into  
 CC the seed's genome on chromosome 8 and near position 117, between markers  
 CC Z1B3 and UMC150a. The maize plants and seeds exhibit resistance to  
 CC Lepidopteran insects since they express the protein CryIAb in high  
 CC levels, and also exhibit resistance to non-selective herbicide  
 CC glufosinate. The transgenic maize is suited for preparation of food  
 CC materials for human or animal consumption e.g. sweet corn for packaging  
 CC or fresh use as a human food, or grain or silage made from field corn,  
 CC containing reduced levels of fungal toxins. The present sequence  
 CC represents a 3' terminator from nopaline synthetase, which is used to  
 CC produce transgenic maize of the invention.  
 XX  
 XX SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;  
 Best Local Similarity 100.0%; Pred. No. 0.085;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21

```

Db 247 GTAACATAGATGACACCGCGC 227
|||||
RESULT 6
AAF89833/C
ID AAF89833 standard; DNA; 249 BP.
XX
AC AAF89833;
XX
DT 23-JUL-2001 (first entry)
XX
DE Nucleotide sequence of a 3' terminator from nopaline synthetase.
XX
KW Maize; Inbred line R660H; 35S constitutive promoter; CryIAb protein;
KW alcohol dehydrogenase; phosphinothricin acetyl transferase; Chromosome 8;
KW Z1B3; UMC150a; Lepidoptera; glufosinate; sweet corn; fungal toxin; ds.
XX
OS Unidentified.
XX
PN US6232534-B1.
XX
PD 15-MAY-2001.
XX
PF 11-JUN-1999; 99US-0330737.
XX
PR 14-MAR-1997; 97US-0818573.
PR 13-MAR-1998; 98US-0042426.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;
XX
DR WPI; 2001-335092/35.
XX
PT Novel seed of maize inbred line R660H, useful for producing inbred
PT maize plants which on crossing with other different maize plants
PT produce hybrid maize seeds and plants having resistance to insects and
PT herbicide -
XX
PS Claim 3; Column 23-24; 24pp; English.
XX
CC The specification describes seed and plants of maize inbred line R660H.
CC The seed comprises two cassettes. The first cassette comprises a
CC cauliflower mosaic virus (CaMV) 35S constitutive promoter operably linked
CC to a maize alcohol dehydrogenase intron, a DNA sequence of a gene
CC encoding a CryIAb protein, and a terminator which functions in plant
CC second cassette comprises a CaMV 35S promoter which functions in plant
CC cells operably linked to a maize alcohol dehydrogenase intron, a DNA
CC sequence of a gene encoding for phosphinothricin acetyl transferase, and
CC a terminator functional in plant. The two cassettes are transcribed in
CC the same direction and the nucleic acid construct is incorporated into
CC the seed's genome on chromosome 8 and near position 117, between markers
CC Z1B3 and UMC150a. The maize plants and seeds exhibit resistance to
CC Lepidopteran insects since they express the protein CryIAb in high
CC levels, and also exhibit resistance to non-selective herbicide
CC glufosinate. The transgenic maize is suited for preparation of food
CC materials for human or animal consumption e.g. sweet corn for packaging
CC or fresh use as a human food or grain or silage made from field corn,
CC containing reduced levels of fungal toxins. The present sequence
CC represents a 3' terminator from nopaline synthetase, which is used to
CC produce transgenic maize of the invention.
XX
SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;
Best Local Similarity 100.0%; Pred. NO. 0.085;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21
|||||
Db 247 GTAACATAGATGACACCGCGC 227

RESULT 8
AAF898689/C
ID AAF898689 standard; DNA; 254 BP.
XX

```

```

RESULT 7
AAZ51175/C
ID AAZ51175 standard; DNA; 253 BP.
XX
AC AAZ51175;
XX
DT 06-JUN-2000 (first entry)
XX
DE 3'UTR of nopaline synthase gene from Agrobacterium tumefaciens.
XX
KW Recombinant expression vector; intron; 5' non-translated leader;
KW fructose-1,6-bisphosphatase; fbp; chlorophyll a/b binding protein;
KW heat shock protein; hsp; peroxidase; beta-tubulin; btub; amylase; actin;
KW sucrose synthase; phenylalanine ammonia lyase; ubiquitin; glutelin;
KW 3' non-translated terminator; lactate dehydrogenase; transgenic plant;
KW resistance; herbicide tolerance; T-DNA; nopaline synthase; 3' nos; ds.
XX
OS Agrobacterium tumefaciens.
XX
PN WO200011200-A2.
XX
PD 02-MAR-2000.
XX
PF 18-AUG-1999; 99WO-US19102.
PR 19-AUG-1998; 98US-0097150.
XX
PA (MONS ) MONSANTO CO.
XX
PI Conner TW, Santino CG;
XX
DR WPI; 2000-237660/20.
XX
PT Novel plants expression vectors, containing combinations of plant
PT introns and non-translated 5' and 3' elements, used for introducing
PT agronomically desirable traits into plants -
XX
PS Example 1; Page 90; 95pp; English.
XX
CC The patent discloses recombinant plant expression vectors, which comprise
CC combinations of introns and 5' and 3' non-translated genetic elements.
CC The vectors comprise a promoter sequence, a 5' non-translated leader
CC sequence (isolated from genes of wheat fructose-1,6-bisphosphatase (fbp),
CC chlorophyll a/b binding protein, heat shock protein (hsp), peroxidase,
CC rice beta-tubulin (btub) or amylase), an intron (isolated from genes of
CC rice actin, sucrose synthase, phenylalanine ammonia lyase, amylase or
CC maize hsp), a DNA coding sequence and a 3' non-translated terminator
CC sequence (isolated from genes of wheat hsp, ubiquitin, fbp, rice
CC glutelin, lactate dehydrogenase or btub). They are used to produce
CC transgenic plants showing resistance to microbial and fungal disease,
CC herbicide tolerance, increased yield and nutritional enhancement. The
CC present sequence is the 3'UTR of nopaline synthase gene (3' nos) of
CC the T-DNA in Agrobacterium tumefaciens. It is used in the construction
CC of recombinant expression vector pMON19433.
XX
SQ Sequence 253 BP; 81 A; 38 C; 48 G; 86 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 253;
Best Local Similarity 100.0%; Pred. NO. 0.085;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21
|||||
Db 247 GTAACATAGATGACACCGCGC 227

RESULT 8
AAF898689/C
ID AAF898689 standard; DNA; 254 BP.
XX

```

```

AC AAZ98689;
XX
DT 06-JUN-2000 (first entry)
XX
DE Nos terminator nucleotide sequence.
XX
KW Endoxylanase; barley; beer fermentation; barley malt; wheat malt;
KW cereal grain malt; mashing; Nos terminator; ds.
XX
OS Synthetic.
XX
PN US6031155-A.
XX
PD 29-FEB-2000.
XX
XX 05-JUN-1997; 97US-0869696.
XX
XX 05-JUN-1997; 97US-0869696.
XX
PA (VDOO/) VAN DEN DOOL R T M.
PA (CAME/) CAMERON-MILLS V.
PA (LOKE/) LOK F.
PA (SINJ/) SINJORGIO C M C.
PA (CASP/) CASPERS M P M.
PA (VALK/) VAN ZEIJL-VAN DER VALK M J.
XX
XX Lok F, Caspers MP, Cameron-Mills V, Van Den Dool RTW;
PI Sinjorgio CMC, Van Zeijl-van der Valk MJ;
PI
XX WPI; 2000-205236/18.
DR
XX Isolated nucleic acid sequence encoding barley endoxylanase, for
PT expressing enhanced amounts of endoxylanase in plant cells, useful in
PT fermentation of beer -
XX
XX Example 5; Column 45-46; 44pp; English.
PS
XX This sequence represents the Nos terminator sequence. The terminator can
CC be used in the production of barley endoxylanase plasmids. The invention
CC relates to a nucleotide sequence encoding the barley endoxylanase
CC protein. The endoxylanase has a molecular weight of 62kD. Endoxylanase is
CC a xylan-degrading enzyme produced by plants for example during
CC germination of cereal grain. Xylan is a constituent of plant cell walls,
CC and its degradation is very important in commercial processes that use
CC cereal grains, such as beer brewing. A nucleic acid construct comprising
CC the endoxylanase nucleotide sequence, a heterologous signal peptide and a
CC promoter, can be used to transform a host cell. Endoxylanase production
CC in a plant can be enhanced through transformation of the plant using the
CC barley endoxylanase nucleotide sequence. The barley endoxylanase
CC nucleotide sequence is useful for expressing enhanced amounts of
CC endoxylanase in plant cells, permitting enhanced degradation of cell wall
CC xylan. Degradation of cell wall xylan is particularly important in
CC fermentation processes that rely on fermentable sugars and nutrients
CC provided by degradation of cereal grains. Barley malt, wheat malt, and
CC cereal grain malt are primary sources of required nutrients in the
CC brewing process. When brewing beer, the amount of starch and protein
CC degradation during malting and mashing greatly impacts the quality of the
CC final product.
XX
SQ Sequence 254 BP; 81 A; 38 C; 49 G; 86 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 254;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21
Db 247 GTAACATAGATGACACCGCGC 227
XXXXXXXXXXXXXXXXXXXX

RESULT 9
AAZ51653/C

```

```

ID AAZ51653 standard; DNA; 257 BP.
XX
AC AAZ51653;
XX
DT 21-JUN-2000 (first entry)
XX
DE Agrobacterium nos transcription termination and polyadenylation sequence.
XX
KW Bacillus thuringiensis; delta-endotoxin; Cry3B; Bt toxin; insect pest;
KW crystal protein; Coleopteran; expression cassette; transgenic plant;
KW insecticide; nos; nopaline synthase gene; transcription termination;
KW polyadenylation; ds.
XX
XX Agrobacterium tumefaciens.
XX
XX WO200011185-A2.
XX
XX 02-MAR-2000.
XX
XX 19-AUG-1999; 99WO-US18883.
XX
XX 19-AUG-1998; 98US-0097150.
XX
XX (MONS ) MONSANTO CO.
XX
XX Romano CP;
PI
XX WPI; 2000-246568/21.
DR
XX Novel expression cassettes which express Bacillus thuringiensis Cry3
PT delta-endotoxin portion which is toxic to coleopteran insect pests,
PT useful for producing transgenic plants with improved insecticidal
PT activity -
XX
XX Claim 14; Page 158; 171pp; English.
PS
XX The present sequence is a Agrobacterium tumefaciens nopaline synthase
CC (nos) transcription termination and polyadenylation sequence
CC which is used in an expression cassette that provides
CC improved expression of Bacillus thuringiensis delta-endotoxin, Cry3B
CC or Cry3B variant proteins, in transgenic plants e.g. maize.
CC Transgenic plants expressing higher levels of Cry3B
CC proteins exhibit increased insecticidal activity against Coleopteran
CC pests.
XX
SQ Sequence 257 BP; 81 A; 41 C; 48 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 257;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21
Db 251 GTAACATAGATGACACCGCGC 231
XXXXXXXXXXXXXXXXXXXX

RESULT 10
AAZ5461/C
ID AAA75461 standard; DNA; 261 BP.
XX
XX AAA75461;
XX
DT 15-JAN-2001 (first entry)
XX
DE Nucleotide sequence of a nopaline synthase terminator.
XX
XX 35S promoter; alcohol dehydrogenase intron; nopaline synthase terminator;
KW insecticidal Cry lab protein toxin; transgenic plant; fungal toxin;
KW aflatoxin; baculovirus; Lepidopteran insect; glufoisinate; ds.
XX
XX Unidentified.
XX

```

PN US6114608-A.  
 XX  
 PD 05-SEP-2000.  
 XX  
 PF 13-MAR-1998; 98US-0042426.  
 XX  
 PR 14-MAR-1997; 97US-0109808.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Dietrich PS, Mettler IJ, Sinibaldi RM;  
 XX  
 DR WPI; 2000-586487/55.  
 XX  
 PT Novel DNA construct comprising an expression cassette having a  
 functional constitutive promoter, operably linked to a maize alcohol  
 dehydrogenase intron, gene encoding Cry 1Ab protein and a terminator  
 XX  
 PS Claim 1; Column 21-22; 24pp; English.  
 XX  
 CC The present sequence represents a nopaline synthase terminator, which is  
 used to create the construct of the invention. The specification  
 describes a nucleic acid construct, comprising an expression cassette  
 containing a cauliflower mosaic virus (CaMV) 35S promoter, a maize  
 alcohol dehydrogenase intron, a DNA molecule encoding an insecticidal  
 Cry 1Ab protein toxin and a NOS (nopaline synthase) terminator in  
 operable linkage. The nucleic acid constructs are useful for  
 producing transgenic plants such as maize, wheat, barley, sorghum and  
 rice. Preferably, maize plants such as field corn, sweet corn, white  
 corn, silage corn or popcorn are produced. The transgenic maize is  
 used for preparing food materials with reduced levels of fungal toxins,  
 e.g. aflatoxins. The transgenic plants express a baculovirus crystal  
 protein toxin, which is effective against Lepidopteran insects at  
 relatively high levels and further provides resistance to the  
 non-selective herbicide glufosinate.  
 XX  
 SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 0.086;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gtaacatagatgacacgcgc 21  
 ||||||||||||||||||  
 DB 253 GTAACATAGATGACACGCGC 233  
 RESULT 11  
 AAD06309/C  
 ID AAD06309 standard; DNA; 261 BP.  
 XX  
 AC AAD06309;  
 XX  
 DT 10-AUG-2001 (first entry)  
 DE  
 XX  
 DE Nopaline synthase terminator of Bt kurstaki expression cassette.  
 XX  
 KW Transgenic maize; expression cassette; 35S promoter;  
 KW alcohol dehydrogenase intron; Cry1Ab toxin; nopaline synthetase;  
 KW NOS terminator; foodstuffs preparation; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN US6222104-B1.  
 XX  
 PD 24-APR-2001.  
 XX  
 PF 13-APR-1999; 99US-0291238.  
 XX  
 PR 09-NOV-1994; 94US-0336627.  
 PR 22-AUG-1996; 96US-0716836.  
 PR 14-MAR-1997; 97US-0818573.

PR 13-MAR-1998; 98US-0042426.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Mettler IJ, Krier M, Mies D;  
 XX  
 DR WPI; 2001-327266/34.  
 XX  
 PT Novel transgenic maize seed for hybrid maize plant production,  
 comprising expression cassette linked operably with CaMV 35S promoter,  
 alcohol dehydrogenase intron, insecticidal Cry1Ab toxin gene or NOS  
 terminator  
 XX  
 PS Claim 1; Column 21-22; 24pp; English.  
 XX  
 CC The present invention relates to transgenic maize seed for hybrid maize  
 plant production, comprising an expression cassette linked operably with  
 constitutive Cauliflower mosaic virus (CaMV) 35S promoter, an alcohol  
 dehydrogenase intron, a DNA encoding an insecticidal  
 Bacillus thuringiensis (Bt) Cry1Ab toxin and a nopaline synthetase (NOS)  
 terminator. The transgenic maize seed is used in maize cultivation  
 methods for propagating hybrid maize seeds and for growing hybrid maize  
 plants. The maize is also used in foodstuffs preparation for animal or  
 human consumption. The inbred maize line NP948 of the invention has  
 reduced level of fungal toxins, hence suitable for foodstuffs  
 preparation. Yield is high. Plant health is improved due to reduced  
 stalk rot. Grain test weight is high and the rate of grain dry down is  
 reduced. The present sequence is nopaline synthetase terminator of Bt  
 kurstaki expression cassette.  
 XX  
 SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 0.086;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gtaacatagatgacacgcgc 21  
 ||||||||||||||||||  
 DB 253 GTAACATAGATGACACGCGC 233  
 RESULT 12  
 AAD06363/C  
 ID AAD06363 standard; DNA; 261 BP.  
 XX  
 AC AAD06363;  
 XX  
 DT 10-AUG-2001 (first entry)  
 DE  
 XX  
 DE 3' nopaline synthase terminator of Btk gene expression cassette.  
 XX  
 KW Maize; expression cassette; 35S promoter; alcohol dehydrogenase intron;  
 KW crystal 1Ab; Cry1Ab toxin; phosphinothricin acetyl transferase; PAT;  
 KW nopaline synthase; NOS terminator; aflatoxin; food material;  
 KW sweet corn; human food; Adhl-1S; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN US6229075-B1.  
 XX  
 PD 08-MAY-2001.  
 XX  
 PF 11-JUN-1999; 99US-0330760.  
 XX  
 PR 14-MAR-1997; 97US-0109808.  
 PR 13-MAR-1998; 98US-0042426.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;  
 XX  
 DR WPI; 2001-342708/36.

XX New maize (Zea mays) inbred line R412H (ATCC 209675), useful for  
 PT producing corn with reduced levels of toxins (e.g. the fungal  
 PT aflatoxin) and for preparing food materials for human or animal  
 PT consumption -  
 XX  
 PS Claim 2; Column 19-20; 30pp; English.  
 XX  
 CC The present invention relates to maize inbred line R412H which comprises  
 CC a nucleic acid construct with two cassettes, which are transcribed in the  
 CC same direction. The expression cassette comprises a Cauliflower mosaic  
 CC virus (CaMV) 35S constitutive promoter operably linked to a maize alcohol  
 CC dehydrogenase intron, a DNA sequence of a gene encoding a  
 CC Bacillus thuringiensis (Bt) crystal IAB (CryIAB) toxin or  
 CC phosphinothricin acetyl transferase (PAT) and a nopaline synthase (NOS)  
 CC terminator functional in plants. The maize inbred line R412H is useful  
 CC for producing corn with reduced levels of toxins (e.g. the fungal  
 CC aflatoxin). This maize line is particularly useful for preparing food  
 CC materials for human or animal consumption, e.g. sweet corn for  
 CC packaging or fresh use as human food, or grain or silage made from field  
 CC corn. The present sequence is 3' NOS terminator of Btk expression  
 CC cassette, which is related to the invention.  
 XX  
 SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 261;  
 Best Local Similarity 100.0%; Pred. NO. 0.086;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
 |||||  
 DB 253 GTAACATAGATGACACCGCGC 233

RESULT 13  
 AAF89820/c  
 ID AAF89820 standard; DNA; 261 BP.

XX AAF89820;  
 AC  
 XX  
 DT 23-JUL-2001 (first entry)  
 XX  
 DE Nucleotide sequence of a 3' terminator from nopaline synthetase.  
 XX  
 KW Maize; Inbred line R660H; 35S constitutive promoter; CryIAB protein;  
 KW alcohol dehydrogenase; phosphinothricin acetyl transferase; chromosome 8;  
 KW Z1B3; UMC150a; Lepidoptera; glufosinate; sweet corn; fungal toxin; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN US6232533-B1.  
 XX  
 PD 15-MAY-2001.  
 XX  
 PF 09-JUN-1999; 99US-0328473.  
 XX  
 PR 14-MAR-1997; 97US-0109808.  
 PR 13-MAR-1998; 98US-0042426.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;  
 XX  
 DR WPI; 2001-335091/35.

XX Novel seed of maize inbred line R372H, useful for producing inbred  
 PT maize plants which on crossing with other different maize plants  
 PT produce hybrid maize seeds and plants having resistance to insects and  
 PT herbicide -  
 XX

PS Claim 2; Column 19-20; 29pp; English.

XX

CC The specification describes seed and plants of maize inbred line R660H.  
 CC The seed comprises two cassettes. The first cassette comprises a  
 CC cauliflower mosaic virus (CaMV) 35S constitutive promoter operably linked  
 CC to a maize alcohol dehydrogenase intron, a DNA sequence of a gene  
 CC encoding a CryIAB protein, and a terminator functional in plants. The  
 CC second cassette comprises a CaMV 35S promoter which functions in plant  
 CC cells operably linked to a maize alcohol dehydrogenase intron, a DNA  
 CC sequence of a gene encoding for phosphinothricin acetyl transferase, and  
 CC a terminator functional in plant. The two cassettes are transcribed in  
 CC the same direction and the nucleic acid construct is incorporated into  
 CC the seed's genome on chromosome 8 and near position 117, between markers  
 CC Z1B3 and UMC150a. The maize plants and seeds exhibit resistance to  
 CC Lepidopteran insects since they express the protein CryIAB in high  
 CC levels, and also exhibit resistance to non-selective herbicide  
 CC glufosinate. The transgenic maize is suited for preparation of food  
 CC materials for human or animal consumption e.g. sweet corn for packaging  
 CC or fresh use as a human food, or grain or silage made from field corn,  
 CC containing reduced levels of fungal toxins. The present sequence  
 CC represents a 3' terminator from nopaline synthetase, which is used to  
 CC produce transgenic maize of the invention.  
 XX  
 SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 261;  
 Best Local Similarity 100.0%; Pred. NO. 0.086;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
 |||||  
 DB 253 GTAACATAGATGACACCGCGC 233

RESULT 14  
 AAF89829/c  
 ID AAF89829 standard; DNA; 261 BP.

XX AAF89829;  
 AC  
 XX  
 DT 23-JUL-2001 (first entry)  
 XX  
 DE Nucleotide sequence of a 3' terminator from nopaline synthetase.  
 XX  
 KW Maize; Inbred line R660H; 35S constitutive promoter; CryIAB protein;  
 KW alcohol dehydrogenase; phosphinothricin acetyl transferase; chromosome 8;  
 KW Z1B3; UMC150a; Lepidoptera; glufosinate; sweet corn; fungal toxin; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN US6232534-B1.  
 XX  
 PD 15-MAY-2001.  
 XX  
 PF 11-JUN-1999; 99US-0330737.  
 XX  
 PR 14-MAR-1997; 97US-0818573.  
 PR 13-MAR-1998; 98US-0042426.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;  
 XX  
 DR WPI; 2001-335092/35.

XX Novel seed of maize inbred line R660H, useful for producing inbred  
 PT maize plants which on crossing with other different maize plants  
 PT produce hybrid maize seeds and plants having resistance to insects and  
 PT herbicide -  
 XX

PS Claim 2; Column 19-20; 24pp; English.

CC The specification describes seed and plants of maize inbred line R660H.  
 CC The seed comprises two cassettes. The first cassette comprises a

CC cauliflower mosaic virus (CaMV) 35S constitutive promoter operably linked  
 CC encoding a CryIAb protein, and a terminator functional in plants. The  
 CC second cassette comprises a CaMV 35S promoter which functions in plant  
 CC cells operably linked to a maize alcohol dehydrogenase intron, a DNA  
 CC sequence of a gene encoding for phosphinothricin acetyl transferase, and  
 CC a terminator functional in plant. The two cassettes are transcribed in  
 CC the same direction and the nucleic acid construct is incorporated into  
 CC the seed's genome on chromosome 8 and near position 117, between markers  
 CC Z1B3 and UMC150a. The maize plants and seeds exhibit resistance to  
 CC Lepidopteran insects since they express the protein CryIAb in high  
 CC levels, and also exhibit resistance to non-selective herbicide  
 CC glufosinate. The transgenic maize is suited for preparation of food  
 CC materials for human or animal consumption e.g. sweet corn for packaging  
 CC or fresh use as a human food, or grain or silage made from field corn,  
 CC containing reduced levels of fungal toxins. The present sequence  
 CC represents a 3' terminator from nopaline synthetase, which is used to  
 CC produce transgenic maize of the invention.  
 XX  
 SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 0.086;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gtaacatagatgacacccgcg 21  
 Db 253 GTACATAGATGACACCGCGC 233  
 |||||

RESULT 15  
 AAT66536/c  
 ID AAT66536 standard; DNA; 263 BP.  
 XX  
 AC AAT66536;  
 XX  
 DN 22-JUL-1997 (first entry)  
 XX  
 DE Nopaline synthase 3' nontranslated region.  
 XX  
 KW Transgenic plant; selectable marker; carotenoid; pigment;  
 KW nopaline synthase; phytoene synthase; plasmid pET0203; ss.  
 XX  
 OS Agrobacterium sp.  
 XX  
 PN WO9714807-A1.  
 XX  
 PD 24-APR-1997.  
 XX  
 PF 29-MAR-1996; 96WO-US04313.  
 XX  
 PR 16-OCT-1995; 95US-0543608.  
 XX  
 PA (SEMI-) SEMINIS VEGATABLES.  
 XX  
 PI Braun CJ, Trulson AJ;  
 XX  
 DR WPI; 1997-245122/22.  
 XX

Visual identification of transgenic plant material - from production  
 of carotenoid pigment encoded by cassette containing Erwinia  
 phytoene synthase gene, useful for selecting material for  
 regeneration

PS Example; Page 37; 62pp; English.  
 PS  
 XX The 3' nontranslated region (AAT66536) of the nopaline synthase  
 CC gene can be cloned from binary vector pBI121 (Clontech). It  
 CC contains sequences that in plant cells result in the termination  
 CC of transcription and additional sequences that when transcribed  
 CC into RNA result in the addition of a poly-A tract to the 3' end  
 CC of the RNA. It is a preferred component of pET0203 (ATCC 97282), a

CC binary vector used in a method for the visual identification of  
 CC transgenic plant material. The vector includes an expression  
 CC cassette comprising the tomato E8 promoter (see also AAT66533), a  
 CC plastid targeting signal (AAT66535) fused to the Erwinia herbicola  
 CC phytoene synthase gene (AAT66534), and the 3' non-translated region.  
 CC Transgenic plant cells and tissues are identified by the  
 CC appearance of orange colour.  
 XX  
 SQ Sequence 263 BP; 83 A; 42 C; 49 G; 89 T; 0 other;  
 Query Match 100.0%; Score 21; DB 18; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 0.086;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gtaacatagatgacacccgcg 21  
 Db 257 GTACATAGATGACACCGCGC 237  
 |||||

Search completed: February 15, 2002, 19:01:09  
 Job time: 20713 sec





**THIS PAGE BLANK (USPTO,**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 15:03:46 ; Search time 353.79 Seconds  
(without alignments)  
13.443 Million cell updates/sec

Title: US-09-698-903B-2

Perfect score: 21

Sequence: 1 gtaacatagatgacacgcgc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2.6/ptodata/2/ina/PCUS\_COMB.seq.\*
- 6: /cgn2.6/ptodata/2/ina/backfiles1.seq.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	21	100.0	249	3	US-09-042-426-8
C 2	21	100.0	249	4	US-09-291-238-8
C 3	21	100.0	249	4	US-09-330-760-8
C 4	21	100.0	249	4	US-09-328-473-8
C 5	21	100.0	249	4	US-09-330-737-8
C 6	21	100.0	254	3	US-08-869-696-18
C 7	21	100.0	261	3	US-09-042-426-4
C 8	21	100.0	261	4	US-09-291-238-4
C 9	21	100.0	261	4	US-09-330-760-4
C 10	21	100.0	261	4	US-09-328-473-4
C 11	21	100.0	261	4	US-09-330-737-4
C 12	21	100.0	1287	1	US-08-064-121-3
C 13	21	100.0	1287	1	US-08-478-015-3
C 14	21	100.0	1287	3	US-08-475-975-3
C 15	21	100.0	1287	3	US-09-084-889-3
C 16	21	100.0	1559	3	US-09-049-475-6
C 17	21	100.0	1829	1	US-07-966-187-17
C 18	21	100.0	1863	1	US-08-525-507-16
C 19	21	100.0	1863	2	US-08-475-427-9
C 20	21	100.0	1863	2	US-07-842-165-9
C 21	21	100.0	2115	2	US-08-767-026-3
C 22	21	100.0	2145	1	US-08-592-936B-16
C 23	21	100.0	2145	2	US-09-111-573-16
C 24	21	100.0	2319	3	US-09-049-475-5
C 25	21	100.0	2345	3	US-09-026-673-1
C 26	21	100.0	2633	1	US-08-452-267-2
C 27	21	100.0	2633	3	US-09-123-644-2

C 28	21	100.0	2917	1	US-08-592-936B-20	Sequence 20, Appl
C 29	21	100.0	2917	2	US-09-111-573-20	Sequence 20, Appl
C 30	21	100.0	3183	1	US-08-413-135-1	Sequence 1, Appl
C 31	21	100.0	3237	2	US-08-419-075-26	Sequence 26, Appl
C 32	21	100.0	3544	2	US-08-485-139-3	Sequence 3, Appl
C 33	21	100.0	3544	2	US-08-485-139-3	Sequence 3, Appl
C 34	21	100.0	3544	3	US-08-750-357-3	Sequence 3, Appl
C 35	21	100.0	3544	3	US-08-750-357-3	Sequence 3, Appl
C 36	21	100.0	3658	3	US-08-894-440-3	Sequence 3, Appl
C 37	21	100.0	3658	3	US-08-894-440-3	Sequence 3, Appl
C 38	21	100.0	4284	1	US-08-525-507-14	Sequence 14, Appl
C 39	21	100.0	4583	3	US-08-810-720-9	Sequence 9, Appl
C 40	21	100.0	4808	1	US-08-351-413-17	Sequence 17, Appl
C 41	21	100.0	4808	1	US-08-351-413-17	Sequence 17, Appl
C 42	21	100.0	4808	2	US-09-025-583-17	Sequence 17, Appl
C 43	21	100.0	4808	2	US-09-025-583-17	Sequence 17, Appl
C 44	21	100.0	4883	1	US-08-064-121-4	Sequence 4, Appl
C 45	21	100.0	4883	3	US-09-084-889-4	Sequence 4, Appl

RESULT 1

US-09-042-426-8/c  
; Sequence 8, Application US/09042426  
; Patent No. 6114608

; GENERAL INFORMATION:

; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi

; TITLE OF INVENTION: DNA Construct Containing Bacillus

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6114608artis Corporation

; STREET: 564 Morris Avenue

; CITY: Summit

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07901

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/042,426

; FILING DATE: March 13, 1998

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Hoxie, Thomas

; REGISTRATION NUMBER: 32,993

; REFERENCE/DOCKET NUMBER: 135/1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (919) 541-8614

; TELEFAX: (919) 541-8689

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 249 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; IMMEDIATE SOURCE:

; CLONE: NOS Terminator

; US-09-042-426-8

Query Match 100.0%; Score 21; DB 3; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gtaacatagatgacacgcgc 21

```

Db      247  GTAACATAGATGACACCGCGC 227
|||||
RESULT 2
US-09-291-238-8/c
; Sequence 8, Application US/09291238
; Patent No. 6222104
; GENERAL INFORMATION:
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
; TITLE OF INVENTION: DNA Construct Containing Bacillus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 622104artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/291,238
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoxie, Thomas
; REGISTRATION NUMBER: 32,993
; REFERENCE/DOCKET NUMBER: 135/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8614
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: NOS Terminator
; US-09-291-238-8

Query Match      100.0%; Score 21; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  gtaacatagatgacacccgc 21
|||||
Db      247  GTAACATAGATGACACCGCGC 227

RESULT 3
US-09-330-760-8/c
; Sequence 8, Application US/09330760
; Patent No. 6229075
; GENERAL INFORMATION:
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
; TITLE OF INVENTION: DNA Construct Containing Bacillus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6229075artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/330,760
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoxie, Thomas
; REGISTRATION NUMBER: 32,993
; REFERENCE/DOCKET NUMBER: 135/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8614
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: NOS Terminator
; US-09-330-760-8

Query Match      100.0%; Score 21; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  gtaacatagatgacacccgc 21
|||||
Db      247  GTAACATAGATGACACCGCGC 227

RESULT 4
US-09-328-473-8/c
; Sequence 8, Application US/09328473
; Patent No. 6232533
; GENERAL INFORMATION:
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
; TITLE OF INVENTION: DNA Construct Containing Bacillus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6232533artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/328,473
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS Terminator  
; US-09-328-473-8

Query Match 100.0%; Score 21; DB 4; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
|||||  
Db 247 GTAACATAGATGACACGCGC 227

RESULT 5  
US-09-330-737-8/c  
; Sequence 8, Application US/09330737  
; Patent No. 6232534  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6232534artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/330,737  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:

; CLONE: NOS Terminator  
; US-09-330-737-8  
  
Query Match 100.0%; Score 21; DB 4; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 gtaacatagatgacacgcgc 21  
|||||  
Db 247 GTAACATAGATGACACGCGC 227  
  
RESULT 6  
US-08-869-696-18/c  
; Sequence 18, Application US/08869696C  
; Patent No. 6031155  
; GENERAL INFORMATION:  
; APPLICANT: Cameron-Mills, Verena  
; APPLICANT: Lok, Finn  
; APPLICANT: Sinjorgo, Catharina Maria Cornelia  
; APPLICANT: Van Den Dool, Ronald Tako Marinus  
; APPLICANT: Caspers, Martinus Petrus Maria  
; APPLICANT: Van Zeijl-van Der Valk, Maria Joanna  
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION  
; FILE REFERENCE: 11225.01US01  
; CURRENT APPLICATION NUMBER: US/08/869,696C  
; CURRENT FILING DATE: 1997-06-05  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 254  
; TYPE: DNA  
; ORGANISM: synthetic  
; US-08-869-696-18

Query Match 100.0%; Score 21; DB 3; Length 254;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
|||||  
Db 247 GTAACATAGATGACACGCGC 227

RESULT 7  
US-09-042-426-4/c  
; Sequence 4, Application US/09042426  
; Patent No. 6114608  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6114608artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993

;; REFERENCE/DOCKET NUMBER: 135/1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (919) 541-8614  
;; TELEFAX: (919) 541-8689  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 261 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; IMMEDIATE SOURCE:  
;; CLONE: NOS terminator  
US-09-042-426-4

Query Match 100.0%; Score 21; DB 3; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 8  
US-09-291-238-4/c  
; Sequence 4, Application US/09291238  
; Patent No. 622104  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 622104artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/291.238  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/042.426  
; FILING DATE: March 13, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 261 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS terminator  
US-09-291-238-4

Query Match 100.0%; Score 21; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 9  
US-09-330-760-4/c  
; Sequence 4, Application US/09330760  
; Patent No. 6229075  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6229075artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/330.760  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/042.426  
; FILING DATE: March 13, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 261 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS terminator  
US-09-330-760-4

Query Match 100.0%; Score 21; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233  
RESULT 10  
US-09-328-473-4/c  
; Sequence 4, Application US/09328473  
; Patent No. 6232533  
; GENERAL INFORMATION:

```
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
; TITLE OF INVENTION: DNA Construct Containing Bacillus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 623253artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/328,473
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoxie, Thomas
; REGISTRATION NUMBER: 32,993
; REFERENCE/DOCKET NUMBER: 135/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8614
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: NOS terminator
;
US-09-328-473-4

Query Match 100.0%; Score 21; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 253 GTRACATAGATGACACGC 233

RESULT 11
US-09-330-737-4/c
; Sequence 4, Application US/09330737
; Patent No. 6232534
; GENERAL INFORMATION:
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
; TITLE OF INVENTION: DNA Construct Containing Bacillus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 623253artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/330,737
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoxie, Thomas
; REGISTRATION NUMBER: 32,993
; REFERENCE/DOCKET NUMBER: 135/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8614
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: NOS terminator
;
US-09-330-737-4

Query Match 100.0%; Score 21; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 253 GTRACATAGATGACACGC 233

RESULT 12
US-08-064-121-3/c
; Sequence 3, Application US/08064121
; Patent No. 5641664
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/064,121
; FILING DATE: 24-MAY-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
```

```

; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
; US-08-064-121-3

```

```

Query Match 100.0%; Score 21; DB 1; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21
|||||
Db 1251 GTAACATAGATGACACGC 1231

```

```

RESULT 13
US-08-478-015-3/c
; Sequence 3, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: Gobel, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylat
; OTHER INFORMATION: site derived from Agrobacterium T-DNA nopaline synthase ge
; US-08-478-015-3

```

```

Query Match 100.0%; Score 21; DB 1; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21
|||||
Db 1251 GTAACATAGATGACACGC 1231

```

```

RESULT 14
US-08-475-975-3/c
; Sequence 3, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: Gobel, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```



```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 50403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
; US-08-475-975-3

Query Match 100.0%; Score 21; DB 3; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
|||||
Db 1251 GTAACATAGATGACACCGCGC 1231

RESULT 15
US-09-084-889-3/c
; Sequence 3, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.

```

```

; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/064,121
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
; US-09-084-889-3

Query Match 100.0%; Score 21; DB 3; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
|||||
Db 1251 GTAACATAGATGACACCGCGC 1231

Search completed: February 15, 2002, 19:07:18
Job time: 14612 sec

```

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 12:14:51 ; Search time 9904.61 Seconds  
(without alignments)  
22.783 Million cell updates/sec

Title: US-09-698-903b-2

Perfect score: 21

Sequence: 1 gtaacatagatgacaccgcgc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	17.8	84.8	887	13	AQ363458
c 2	17.4	82.9	300	10	AU114795
c 3	17.4	82.9	833	11	BF104658
c 4	17.4	82.9	936	13	CNS018Y4
5	17	81.0	215	10	BB143973
6	17	81.0	277	10	BB244247
7	16.8	80.0	232	10	AV008164
8	16.8	80.0	379	10	AV684098
9	16.8	80.0	542	13	CNS00WTL
10	16.8	80.0	594	10	AI109780
11	16.8	80.0	662	13	AZ029768
12	16.8	80.0	710	10	BE038757

13	16.8	80.0	826	10	AW940028
14	16.8	80.0	910	11	BG823293
c 15	16.4	78.1	245	10	AA359059
16	16.4	78.1	282	10	BB383808
c 17	16.4	78.1	291	10	AA361337
18	16.4	78.1	299	10	AI081942
19	16.4	78.1	343	10	BE118822
20	16.4	78.1	348	11	BI013588
c 21	16.4	78.1	411	11	N71737
c 22	16.4	78.1	423	11	R58585
c 23	16.4	78.1	435	10	BE710208
24	16.4	78.1	442	10	AV617102
c 25	16.4	78.1	442	10	AV617103
c 26	16.4	78.1	455	11	W92829
c 27	16.4	78.1	460	10	AW652485
c 28	16.4	78.1	470	10	AL725000
c 29	16.4	78.1	478	10	AL121482
c 30	16.4	78.1	494	10	AA010698
c 31	16.4	78.1	577	11	BF342064
c 32	16.4	78.1	612	11	BF038038
c 33	16.4	78.1	632	10	AW607872
c 34	16.4	78.1	643	11	BG705748
c 35	16.4	78.1	657	10	AW853029
c 36	16.4	78.1	692	11	BG702664
c 37	16.4	78.1	693	11	BG422864
c 38	16.4	78.1	699	11	BG714597
c 39	16.4	78.1	704	11	BG747014
c 40	16.4	78.1	742	10	BE409302
c 41	16.4	78.1	747	11	BG325494
c 42	16.4	78.1	750	11	BG701119
c 43	16.4	78.1	802	10	AL546504
c 44	16.4	78.1	834	11	BG716798
c 45	16.4	78.1	851	10	AU117875

#### ALIGNMENTS

RESULT 1  
AQ363458/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AQ363458 887 bp DNA GSS 16-DEC-1999  
nbxb0059A08f CUGI Rice BAC Library Oryza sativa genomic clone  
nbxb0059A08f, DNA sequence.  
AQ363458  
AQ363458.2 GI:6583332  
GSS.  
Oryza sativa.  
Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoidae; Oryzaceae; Oryza.  
1 (bases 1 to 887)  
Wing, R.A. and Dean, R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
On Dec 15, 1999 this sequence version replaced gi:4213113.  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Email: rwing@clemson.edu  
Fax: 864 656 4293  
Seq primer: TAATACGACTCACTATAGGG  
Class: BAC ends  
High quality sequence stop: 1.  
Location/Qualifiers  
1. .887  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbxb0059A08f"

```

/clone_lib="CUCI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBelOAGc1; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."
                                     260 a   200 c   189 g   236 t      2 others
BASE COUNT
ORIGIN

Query Match          84.8%; Score 17.8; DB 13; Length 887;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgc 21
    | ||||| ||||| |||||
Db 517 GCACATTGATGCACCGCGC 497

RESULT 2
AUI14795/c
LOCUS AUI14795 300 bp mRNA EST 19-OCT-2000
DEFINITION AUI14795 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone yk728el0 3', mRNA sequence.
ACCESSION AUI14795
VERSION AUI14795.1 GI:10928362
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
AUTHORS 1 (bases 1 to 300)
TITLE /Rhabditidae; Peioderinae; Caenorhabditis.
COMMENT Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y. and Sugano,S.
A complementary view of the C. elegans genome
Unpublished (2000)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: yokohara@lab.nig.ac.jp.
Location/Qualifiers
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk728el0"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 96 a 49 c 66 g 86 t 3 others
ORIGIN

Query Match          82.9%; Score 17.4; DB 11; Length 833;
Best Local Similarity 94.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taacatatgatgacaccgc 20
    | ||||| ||||| |||||
Db 600 TAACATAGATGACCGCGC 582

RESULT 4
CNS01E4
LOCUS CNS01E4 936 bp DNA GSS 30-MAY-2001
DEFINITION Anopheles gambiae GSS T7 end of clone 03E09 of Notredame1 library from strain PEST of Anopheles gambiae African malaria mosquito),

```

genomic survey sequence.  
AL141149  
AL141149.1 GI:6999267  
GSS.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
Culicoidae; Anopheles.  
1 (bases 1 to 936)  
Genoscope.  
Direct Submission  
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
2 (bases 1 to 936)  
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissensbach,J.  
Direct Submission  
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.  
Roux, Paris 75015, France  
This clone is from an A. gambiae BAC library provided by F.H.  
Collins and sequenced by Genoscope in collaboration with the  
Laboratory of Biochem. and Biol. Molec. of Insects, Institut  
Pasteur.

FEATURES  
source  
1..936  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="03E09"  
/clone\_lib="NotreDamel"  
/note="end : T7"  
267 a 227 c 172 g 263 t 7 others

BASE COUNT 267 a 227 c 172 g 263 t 7 others  
ORIGIN

Query Match 82.9%; Score 17.4; DB 13; Length 936;  
Best Local Similarity 94.7%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aacatagatgacacgcgc 21  
||||||| |||||||  
Db 68 AACATAGAAGACACCGCG 86

RESULT 5  
BB143973 215 bp mRNA EST 28-JUN-2000  
LOCUS BB143973 RIKEN full-length enriched, adult female vagina Mus  
DEFINITION musculus cDNA clone 9930023C03 3', mRNA sequence.  
BB143973  
ACCESSION BB143973.1 GI:8798910  
VERSION  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 215)  
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,  
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,  
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
,Y., Shigenoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,  
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya  
,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaoka,I.,  
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki

TITLE  
COMMENT

Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermotabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Tomaru,Y., Carninci,P., Akiyama,J., Shibata,K., Ozawa,Y., Muramatsu,M., Okazaki  
,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.

FEATURES  
source  
Location/Qualifiers  
1..215  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="9930023C03"  
/clone\_lib="RIKEN full-length enriched, adult female  
vagina"  
/sex="female"  
/tissue\_type="vagina"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAAGATCCAGAGCTCTTTTTTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 185.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATCTCGAGTTAAATTAATATCCCTCCCTCCCTCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I."

BASE COUNT 56 a 41 c 55 g 63 t  
ORIGIN

Query Match 81.0%; Score 17; DB 10; Length 215;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 taacatagatgacacgc 18  
||||||| |||||||  
Db 36 TAACATAGATGACACGC 52

RESULT 6  
BB244247 277 bp mRNA EST 06-JUL-2000  
LOCUS BB244247 RIKEN full-length enriched, 7 days neonate cerebellum Mus  
DEFINITION musculus cDNA clone A730001C13 3', mRNA sequence.  
BB244247  
ACCESSION BB244247.1 GI:8936993  
VERSION  
KEYWORDS EST.  
SOURCE house mouse.



/db_xref="taxon:3702"	
/clone_lib="TAMU"	
/clone="T12G21"	
/note="end : T7"	
BASE COUNT	205 a 88 c 110 g 139 t
ORIGIN	
Query Match 80.0%; Score 16.8; DB 13; Length 542;	
Best Local Similarity 90.0%; Pred. No. 3.9e+02;	
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 gtaacatagatgacaccgcg 20
Db	
518 GTAACATAGATTACACGGCG 537	
RESULT 10	
AL109780	
LOCUS	AL109780 594 bp mRNA EST 23-APR-2001
DEFINITION	GH09125.5prime GH Drosophila melanogaster head pOT2 Drosophila melanogaster cDNA clone GH09125 5 similar to CG16904: FBan0016904 located on: 3R 85E13-85E13;; 04/10/2001, mRNA sequence.
ACCESSION	AL109780
VERSION	AL109780.2 GI:13757702
KEYWORDS	EST.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
TITLE	BDGP/HMI Drosophila EST Project
JOURNAL	Unpublished (2001)
COMMENT	On Aug 26, 1998 this sequence version replaced gi:3478104. Other_ESTs: GH09125.3prime Contact: Stapleton, M. BDGP
Lawrence Berkeley National Lab	
One Cyclotron Rd, Berkeley, CA 94720, USA	
Fax: 510 486 6798	
Email: <a href="http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu">http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu</a>	
hit genomic AE003684: arm:3R [5513047..5732625]	
estimated-cyto:85E6-85F2: 04/10/2001	
Plate: GH.91 row: C column: 1	
High quality sequence stop: 503	
POLYA-No.	
FEATURES	Location/Qualifiers
source	1. .594
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone="GH09125"
	/clone_lib="GH Drosophila melanogaster head pOT2"
	/sex="male and female"
	/dev_stage="adult"
	/lab_host="DH5 - alpha"
	/note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."
BASE COUNT	181 a 129 c 145 g 139 t
ORIGIN	
Query Match 80.0%; Score 16.8; DB 10; Length 594;	
Best Local Similarity 90.0%; Pred. No. 4e+02;	
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
Qy	2 taacatagatgacaccgcg 21
Db	
273 TGACATAGATGACACCGCG 292	





```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH01164"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/notes="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT      218 a   208 c   212 g   188 t
ORIGIN

```

```

Query Match      80.0%; Score 16.8; DB 10; Length 826;
Best Local Similarity 90.0%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY  2 taacatagatgacaccgcgc 21
| | | | | | | | | | | | | | | |
Db  182 TGACATAGATGACACCCCGC 201

```

## RESULT 14

BG823293

## LOCUS

```

DEFINITION      602726676F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4866115 5',
mRNA sequence.

```

## ACCESSION

BG823293

## VERSION

BG823293.1 GI:14170880

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

## REFERENCE

1 (bases 1 to 910)

## AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

```

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1732 row: h column: 20
High quality sequence stop: 649.

```

## FEATURES

## source

1..910

Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4866115"
/clone_lib="NIH_MGC.15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      258 a   203 c   248 g   201 t
ORIGIN

```

## Query Match

80.0%; Score 16.8; DB 11; Length 910;

## Best Local Similarity

90.0%; Pred. No. 4.6e+02;

## Matches 18;

Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY  2 taacatagatgacaccgcgc 21
| | | | | | | | | | | | | | | |
Db  487 TAACATAGATGACACTGCAC 506

```

## RESULT 15

AA359059/c

## LOCUS

```

DEFINITION      EST68007 Fetal lung II Homo sapiens cDNA 5' end similar to similar
to peripheral myelin protein 22, mRNA sequence.

```

## ACCESSION

AA359059

## VERSION

AA359059.1 GI:2011376

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

## REFERENCE

1 (bases 1 to 245)

## AUTHORS

```

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
,M.R., Rosen,C.A., Haseeltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.

```

## TITLE

```

Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence

```

## JOURNAL

Nature 377 (6547 Suppl.), 3-174 (1995)

## MEDLINE

96026280

## COMMENT

```

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavetigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

```

## FEATURES

## source

1..245

Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="ATCC (inhost):163151"
/db_xref="taxon:9606"
/clone_lib="Fetal lung II"
/sex="male"
/dev_stage="fetus, 19 wks"
/notes="Organ: lung; Vector: phuescript SK-; Site_1: EcoRI
; Site_2: XhoI"
BASE COUNT      38 a   74 c   61 g   72 t
ORIGIN

```

## Query Match

78.1%; Score 16.4; DB 10; Length 245;

## Best Local Similarity

94.4%; Pred. No. 5e+02;

## Matches 17;

Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taacatagatgacaccgc 19

| | | | | | | | | | | | | | | |

Db 232 TCACATAGATGACACCGC 215

Tue Feb 19 10:58:37 2002

us-09-698-903b-2.rst

Page 8

Search completed: February 15, 2002, 18:03:38  
Job time: 20927 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:46:25 ; Search time 2553.1 Seconds  
(without alignments)  
135.694 Million cell updates/sec

Title: US-09-698-903B-3

Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_ba:\*
  - 2: gb\_hgt:\*
  - 3: gb\_in:\*
  - 4: gb\_om:\*
  - 5: gb\_ov:\*
  - 6: gb\_pat:\*
  - 7: gb\_ph:\*
  - 8: gb\_pl:\*
  - 9: gb\_pr:\*
  - 10: gb\_ro:\*
  - 11: gb\_sts:\*
  - 12: gb\_sy:\*
  - 13: gb\_un:\*
  - 14: gb\_vl:\*
  - 15: em\_ba:\*
  - 16: em\_fun:\*
  - 17: em\_hum:\*
  - 18: em\_in:\*
  - 19: em\_om:\*
  - 20: em\_or:\*
  - 21: em\_ov:\*
  - 22: em\_pat:\*
  - 23: em\_ph:\*
  - 24: em\_pl:\*
  - 25: em\_ro:\*
  - 26: em\_sts:\*
  - 27: em\_sv:\*
  - 28: em\_un:\*
  - 29: em\_vl:\*
  - 30: em\_htgo\_hum:\*
  - 31: em\_htgo\_inv:\*
  - 32: em\_htgo\_rod:\*
  - 33: em\_htg\_hum:\*
  - 34: em\_htg\_inv:\*
  - 35: em\_htg\_rod:\*
  - 36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	6	AX127750
2	21	100.0	3266	6	AX08065
3	21	100.0	3266	6	AX10061
4	21	100.0	4832	6	AX172441
5	21	100.0	4946	6	A60108
6	21	100.0	4946	6	A76915
7	21	100.0	4946	6	AR098307
8	21	100.0	4946	6	AX172440
9	21	100.0	5865	6	AX127748
10	21	100.0	6254	8	NTR29
11	19.4	92.4	3243	5	XL070985
12	19.4	92.4	149710	2	AC022563
13	19.4	92.4	179183	2	AC019254
14	17.8	84.8	167409	2	AC073374
15	17.8	84.8	179510	2	AC013820
16	17.4	82.9	39443	9	AC010505
17	17.4	82.9	172246	6	AC024941
18	17.4	82.9	195516	2	AC053540
19	17.4	82.9	207856	2	AC078778
20	17	81.0	4900	5	XLAPICAL
21	17	81.0	123016	2	AF189001
22	17	81.0	159934	9	AL160273
23	17	81.0	182942	2	AC087793
24	17	81.0	186413	2	AL161913
25	16.8	80.0	1485	14	AF195032
26	16.8	80.0	1485	14	AF195033
27	16.8	80.0	1485	14	AF195034
28	16.8	80.0	1485	14	AF195035
29	16.8	80.0	1485	14	AF195036
30	16.8	80.0	1485	14	AF195037
31	16.8	80.0	1485	14	AF195038
32	16.8	80.0	1485	14	AF195039
33	16.8	80.0	1485	14	AF264053
34	16.8	80.0	1485	14	DENENVB
35	16.8	80.0	1485	14	DENENVC
36	16.8	80.0	1485	14	DENENVD
37	16.8	80.0	1485	14	DVU34933
38	16.8	80.0	1485	14	DVU34934
39	16.8	80.0	1485	14	DVU34935
40	16.8	80.0	1485	14	DVU34936
41	16.8	80.0	1485	14	DVU34937
42	16.8	80.0	1485	14	DVU34938
43	16.8	80.0	1485	14	DVU34939
44	16.8	80.0	1485	14	DVU34940
45	16.8	80.0	1485	14	TOGDENN2

ALIGNMENTS

RESULT 1

AX127750

LOCUS AX127750 21 bp DNA

DEFINITION Sequence 3 from Patent WO0131042.

ACCESSION AX127750

VERSION AX127750.1 GI:14134397

KEYWORDS .

SOURCE synthetic construct.

ORGANISM synthetic construct.

REFERENCE 1 (bases 1 to 21)

AUTHORS Weston,B. and de Beuckeleer,M.

TITLE Male-sterile brassica plants and methods for producing same

JOURNAL Patent: WO 0131042-A 3 03-MAY-2001;

FEATURES

source

1..21

Location/Qualifiers

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer MLD008"

15-MAY-2001

```

BASE COUNT      4 a      1 c      10 g      6 t
ORIGIN

Query Match      100.0%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 1 ATAGGGTGGAGGCTATTGG 21

RESULT 2
A08065 3266 bp DNA 12-AUG-1993
LOCUS Synthetic nucleotide sequence of the TA29 gene.
DEFINITION
ACCESSION A08065
VERSION A08065.1 GI:413313
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 3266)
AUTHORS
JOURNAL
JOURNAL
FEATURES
Source
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
1527..2492
/gene="TA29"
1527..2492
/gene="TA29"
/codon_start=1
/protein_id="CAA00738.1"
/db_xref="GI:413314"
/translation="MVAPKWFVIFSMILLSLAICSGQPVTSDAIKAKEADHNLKAHT
LSNDKAGFGGGGFGIGGAGGCGGGGSDTNYGNYPCGSIHGCTVPGFGL
PKPVFGVYSPGGCYVCPADIPGTGGMTESKITGISQARLYRCKPGPNMCDKDCNE
LLHFVPMQDKHDKHLYRGGRRIGLTVGGVGGFIFGFGAGGCGGGGSDA
PGCSNDGCDPGCGPGCGYACAPANNPSGGITFPHISLSRFDGPyRCRPMCESEDC
NELLHFVSPMQKHNRHDIHVERSEEAHHQSKHKDEDIIN"

BASE COUNT      1016 a      581 c      623 g      1046 t
ORIGIN

Query Match      100.0%; Score 21; DB 6; Length 3266;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 178 ATAGGGTGGAGGCTATTGG 198

RESULT 4
AX172441/c 4832 bp DNA 03-JUL-2001
LOCUS Sequence 2 from Patent WO0141558.
DEFINITION
ACCESSION AX172441
VERSION AX172441.1 GI:14597553
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 4832)
AUTHORS de Both, G. and de Beuckeleer, M.
TITLE Hybrid winter oilseed rape and methods for producing same
JOURNAL Patent: WO 0141558-A 2 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES
Source
Location/Qualifiers
1..4832
/organism="synthetic construct"
/db_xref="taxon:32630"
/notes="T-DNA of plasmid pTHW118"
1883..4065
/misc_feature
/notes="HpaI restriction fragment"
BASE COUNT      1528 a      883 c      932 g      1488 t
ORIGIN 1 others

Query Match      100.0%; Score 21; DB 6; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 4602 ATAGGGTGGAGGCTATTGG 4582

RESULT 5
A60108/c 4946 bp DNA 06-MAR-1998
LOCUS Sequence 1 from Patent WO9706267.
DEFINITION
ACCESSION A60108
VERSION A60108.1 GI:3715124
```

```
KEYWORDS
SOURCE      Transformation vector pTHW107.
ORGANISM    Transformation vector pTHW107
            artificial sequence; vectors.
REFERENCE   1 (bases 1 to 4946)
AUTHORS    De,B.M.
TITLE      Genetic transformation using a PARP inhibitor
JOURNAL    Patent: WO 9706267-A 1 20-FEB-1997;
            PLANT GENETIC SYSTEMS NV (BE)
FEATURES   Location/Qualifiers
            source
            1..4946
            /organism="Transformation vector pTHW107"
            /db_xref="taxon:126810"
BASE COUNT 1569 a 891 c 963 g 1523 t
ORIGIN

Query Match      100.0%; Score 21; DB 6; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 6
A76915/C
LOCUS      A76915 4946 bp DNA PAT 19-OCT-1999
DEFINITION Sequence 1 from Patent EP0757102.
ACCESSION A76915
VERSION A76915.1 GI:6088712
KEYWORDS   Transformation vector pTHW107.
ORGANISM   Transformation vector pTHW107
            artificial sequence; vectors.
REFERENCE   1 (bases 1 to 4946)
AUTHORS    De,B.M.
TITLE      GENETIC TRANSFORMATION USING A PARP INHIBITOR
JOURNAL    Patent: EP 0757102-A 1 05-FEB-1997;
            PLANT GENETIC SYSTEMS NV (BE)
FEATURES   Location/Qualifiers
            source
            1..4946
            /organism="Transformation vector pTHW107"
            /db_xref="taxon:126810"
BASE COUNT 1569 a 891 c 963 g 1523 t
ORIGIN

Query Match      100.0%; Score 21; DB 6; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 7
AR098307/c
LOCUS      AR098307 4946 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6074876.
ACCESSION AR098307
VERSION AR098307.1 GI:12807564
KEYWORDS   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 4946)
AUTHORS    De Block,M.
TITLE      Genetic transformation using a PARP inhibitor
JOURNAL    Patent: US 6074876-A 1 13-JUN-2000;
            Location/Qualifiers
FEATURES   Location/Qualifiers
            source
            1..4946
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="T-DNA of plasmid pCOL13"
BASE COUNT 1849 a 1095 c 1149 g 1772 t
ORIGIN

KEYWORDS
SOURCE      1..4946
            /organism="unknown"
ORGANISM    1569 a 891 c 963 g 1523 t
REFERENCE   1 (bases 1 to 4946)
AUTHORS    de Both,G. and de Beuckeleer,M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL    Patent: WO 0141558-A 1 14-JUN-2001;
            Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source
            1..4946
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="T-DNA of plasmid pTHW107"
            /misc_feature 964..4906
            /note="Hind III fragment"
BASE COUNT 1569 a 891 c 963 g 1523 t
ORIGIN

Query Match      100.0%; Score 21; DB 6; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 8
AX172440/c
LOCUS      AX172440 4946 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 1 from Patent WO0141558.
ACCESSION AX172440
VERSION AX172440.1 GI:14597552
KEYWORDS   synthetic construct.
ORGANISM    synthetic construct
            artificial sequence.
REFERENCE   1 (bases 1 to 4946)
AUTHORS    de Both,G. and de Beuckeleer,M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL    Patent: WO 0141558-A 1 14-JUN-2001;
            Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source
            1..4946
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="T-DNA of plasmid pTHW107"
            /misc_feature 964..4906
            /note="Hind III fragment"
BASE COUNT 1569 a 891 c 963 g 1523 t
ORIGIN

Query Match      100.0%; Score 21; DB 6; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 9
AX127748/c
LOCUS      AX127748 5865 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 1 from Patent WO0131042.
ACCESSION AX127748
VERSION AX127748.1 GI:14134395
KEYWORDS   synthetic construct.
ORGANISM    synthetic construct
            artificial sequence.
REFERENCE   1 (bases 1 to 5865)
AUTHORS    Weston,B. and de Beuckeleer,M.
TITLE      Male-sterile brassica plants and methods for producing same
JOURNAL    Patent: WO 0131042-A 1 03-MAY-2001;
            Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source
            1..5865
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="T-DNA of plasmid pCOL13"
BASE COUNT 1849 a 1095 c 1149 g 1772 t
ORIGIN
```

```

Query Match      100.0%; Score 21; DB 6; Length 5865;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 4717 ATAGGTGGGAGCTATTGG 4697

RESULT 10
NTA29      NTA29      6254 bp      DNA      PLN      09-MAY-1995
LOCUS      Tobacco anther-specific gene TA-29 and stem-specific gene TSJTL.
DEFINITION X52283
ACCESSION  X52283.1 GI:20035
VERSION    X52283.1
KEYWORDS   tissue specific gene.
SOURCE     common tobacco.
ORGANISM   Nicotiana tabacum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE  1 (bases 1 to 6254)
AUTHORS    Mariani T.
TITLE      Direct Submission
JOURNAL    Submitted (23-MAR-1990) Mariani T., Plant Genetic Systems, J
           Plateaustraat 22, B 9000 Gent, Belgium
AUTHORS    2 (bases 1 to 6254)
TITLE      Seurinck, J., Truettner, J. and Goldberg, R.B.
JOURNAL    The nucleotide sequence of an anther-specific gene
MEDLINE    Nucleic Acids Res. 18 (11), 3403 (1990)
COMMENT    90287727
FEATURES   Data kindly reviewed (13-DEC-1990) by Mariani C.
           Location/Qualifiers
           source
             1..6254
               /organism="Nicotiana tabacum"
               /cultivar="Samsun"
               /db_xref="taxon:4097"
               /clone_lib="lambda Charon32"
               /date="1446..1452"
               /note="TSJTL"
               /gene="TA29"
               /protein_id="CAA36524.1"
               /db_xref="GI:20036"
               /db_xref="SWISS-PROT:P24804"
               /translation="MVAPKVFISFMILLSLAICGQPVTSDAIKAKEADHNLKAHT
               LSNIDAKFGGGGGGIGGGWAGGGGSDTPNYGNYPCSTHGCTVPGFGL
               PKVGVVPSVGGYVCPADIPGGMPTESKITGISARLYRCKPGNMCDSKDCNE
               LLLHFVPMQKHQKHLRYGRRIGLTVGGVGFGLGAGGGGGGGGSDA
               PGCSNDCDPCGCPGCGYACAPANNPSSGITEPHISGLSRFDGPRCPRDMCESED
               NELLHFVSPMQKHENRHHDIHVERDEEAAHQSKQHKDEDIIN"
               /number=3
               /gene="TA29"
               /product="anther-specific"
               /complement(join(<3022..>4678)
               /gene="TSJTL"
               /note="stem specific, weakly expressed in other organs"
               /complement(3022..3075)
               /gene="TSJTL"
               /number=3
               /complement(join(<3022..3075,4047..4287,4473..>4678))
               /gene="TSJTL"
               /note="stem specific, weakly expressed in other organs"
               /complement(3022..4678)
               /gene="TSJTL"

TATA_signal
gene
prim_transcript 1477..(2617,2714)
CDS
XLU70985      3243 bp      DNA      VRT      06-APR-1998
LOCUS      Xenopus laevis transforming growth factor-beta 5 gene, 5' flanking
DEFINITION sequence.
ACCESSION  U70985
VERSION    U70985.1 GI:2267223
KEYWORDS   African clawed frog.
SOURCE     Xenopus laevis
ORGANISM   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
           Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 3243)
AUTHORS    Vempati, U.D. and Kondaliah, P.
TITLE      Characterization of the 5' flanking region of the Xenopus laevis
           transforming growth factor-beta 5 (TGF-beta 5) gene
JOURNAL    Gene 208 (2), 323-329 (1998)
MEDLINE    98201629
REFERENCE  2 (bases 1 to 3243)
AUTHORS    Vempati, U.D. and Kondaliah, P.
TITLE      Direct Submission
JOURNAL    Submitted (17-SEP-1996) Center for Reproductive Biology and
           Molecular Endocrinology, Indian Institute of Science, Bangalore,
           Karnataka 560 012, India
COMMENT    On Jul 21, 1997 this sequence version replaced gi:1809300.
FEATURES   location/Qualifiers
           source
             1..3243
               /organism="Xenopus laevis"
               /db_xref="taxon:8355"
               /number=1
               /note="5' flanking sequence of transforming growth
               factor-beta 5"

mat_peptide
prim_transcript complement(join(<3022..>4678)
exon
misc_feature 1..2890
gene
           /note="5' flanking sequence of transforming growth
           factor-beta 5"

```

## exon

<1..3224  
/number=1  
CDS  
/product="transforming growth factor-beta 5 (TGF-beta 5)"  
/protein\_id="AAC60371.1"  
/db\_xref="GI:2281072"  
/translation="MEVLMMLLVLLVHLSSLAWSLSTCKAVDMEEYKRRKRIEAIKRG  
ILSKLKDKTPDVPDSEKMTVPSEAFILYNSTLEVIKREKATREEHVGHQDQNIQDYAK  
QVYRFESIT"  
3225..>3243  
/number=1

## intron

BASE COUNT 970 a 675 c 676 g 922 t  
ORIGIN

## Query Match

Best Local Similarity 92.4%; Score 19.4; DB 5; Length 3243;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atagggtggaggctatttgg 21

|||||||  
Db 1731 ATAGGGTGGAGGCTATATGG 1711

## RESULT 12

AC022563/c LOCUS AC022563 149710 bp DNA HTG 13-JUL-2000  
DEFINITION Homo sapiens clone RP11-3L23, LOW-PASS SEQUENCE SAMPLING.

## ACCESSION

AC022563

## VERSION

AC022563.2 GI:9121084

## KEYWORDS

HTG; HTGS\_PHASE0.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 149710)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliviar,T.M., Peterson,K.,  
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

## TITLE

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Jul 13, 2000 this sequence version replaced gi:6910806.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2771

Center clone name: 3\_L\_23

-----  
\* NOTE: This record contains 171 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 682: contig of 682 bp in length  
\* 683 782: gap of 100 bp  
\* 783 1492: contig of 710 bp in length  
\* 1493 1592: gap of 100 bp  
\* 1593 2305: contig of 713 bp in length  
\* 2306 2405: gap of 100 bp  
\* 2406 3113: contig of 708 bp in length  
\* 3114 3213: gap of 100 bp  
\* 3214 3867: contig of 654 bp in length  
\* 3868 3967: gap of 100 bp  
\* 3968 4646: contig of 679 bp in length  
\* 4647 4746: gap of 100 bp  
\* 4747 5416: contig of 670 bp in length  
\* 5417 5516: gap of 100 bp  
\* 5517 6193: contig of 677 bp in length  
\* 6194 6293: gap of 100 bp  
\* 6294 6968: contig of 675 bp in length  
\* 6969 7068: gap of 100 bp  
\* 7069 7757: contig of 689 bp in length  
\* 7758 7857: gap of 100 bp  
\* 7858 8557: contig of 700 bp in length  
\* 8558 8657: gap of 100 bp  
\* 8658 9336: contig of 679 bp in length  
\* 9337 9436: gap of 100 bp  
\* 9437 10124: contig of 688 bp in length  
\* 10125 10224: gap of 100 bp  
\* 10225 10879: contig of 655 bp in length  
\* 10880 10979: gap of 100 bp  
\* 10980 11663: contig of 684 bp in length  
\* 11664 11763: gap of 100 bp  
\* 11764 12433: contig of 670 bp in length  
\* 12434 12533: gap of 100 bp  
\* 12534 13209: contig of 676 bp in length  
\* 13210 13309: gap of 100 bp  
\* 13310 13989: contig of 680 bp in length  
\* 13990 14089: gap of 100 bp  
\* 14090 14798: contig of 709 bp in length  
\* 14799 14898: gap of 100 bp  
\* 14899 15571: contig of 673 bp in length  
\* 15572 15671: gap of 100 bp  
\* 15672 16387: contig of 716 bp in length  
\* 16388 16487: gap of 100 bp  
\* 16488 17198: contig of 711 bp in length  
\* 17199 17298: gap of 100 bp  
\* 17299 18008: contig of 710 bp in length  
\* 18009 18108: gap of 100 bp  
\* 18109 18772: contig of 664 bp in length  
\* 18773 18872: gap of 100 bp  
\* 18873 19555: contig of 683 bp in length  
\* 19556 19655: gap of 100 bp  
\* 19656 20320: contig of 665 bp in length  
\* 20321 20420: gap of 100 bp  
\* 20421 21112: contig of 692 bp in length  
\* 21113 21212: gap of 100 bp  
\* 21213 21885: contig of 673 bp in length  
\* 21886 21985: gap of 100 bp  
\* 21986 22671: contig of 686 bp in length  
\* 22672 22771: gap of 100 bp  
\* 22772 23451: contig of 680 bp in length  
\* 23452 23551: gap of 100 bp  
\* 23552 24267: contig of 716 bp in length  
\* \*

\* 24268 24367: gap of 100 bp  
 \* 24368 25073: contig of 706 bp in length  
 \* 25074 25173: gap of 100 bp  
 \* 25174 25867: contig of 694 bp in length  
 \* 25868 25967: gap of 100 bp  
 \* 25968 26669: contig of 702 bp in length  
 \* 26670 26769: gap of 100 bp  
 \* 26770 27445: contig of 676 bp in length  
 \* 27446 27545: gap of 100 bp  
 \* 27546 28224: contig of 679 bp in length  
 \* 28225 28324: gap of 100 bp  
 \* 28325 29002: contig of 678 bp in length  
 \* 29003 29102: gap of 100 bp  
 \* 29103 29785: contig of 683 bp in length  
 \* 29786 29885: gap of 100 bp  
 \* 29886 30571: contig of 686 bp in length  
 \* 30572 30671: gap of 100 bp  
 \* 30672 31357: contig of 686 bp in length  
 \* 31358 31457: gap of 100 bp  
 \* 31458 32168: contig of 711 bp in length  
 \* 32169 32268: gap of 100 bp  
 \* 32269 32951: contig of 683 bp in length  
 \* 32952 33051: gap of 100 bp  
 \* 33052 33756: contig of 705 bp in length  
 \* 33757 33856: gap of 100 bp  
 \* 33857 34568: contig of 710 bp in length  
 \* 34567 34668: gap of 100 bp  
 \* 34667 35376: contig of 710 bp in length  
 \* 35377 35476: gap of 100 bp  
 \* 35477 36143: contig of 667 bp in length  
 \* 36144 36243: gap of 100 bp  
 \* 36244 36924: contig of 681 bp in length  
 \* 36925 37024: gap of 100 bp  
 \* 37025 37701: contig of 677 bp in length  
 \* 37702 37801: gap of 100 bp  
 \* 37802 38493: contig of 692 bp in length  
 \* 38494 38593: gap of 100 bp  
 \* 38594 39274: contig of 681 bp in length  
 \* 39275 39374: gap of 100 bp  
 \* 39375 40071: contig of 697 bp in length  
 \* 40072 40171: gap of 100 bp  
 \* 40172 40870: contig of 699 bp in length  
 \* 40871 40970: gap of 100 bp  
 \* 40971 41659: contig of 689 bp in length  
 \* 41660 41759: gap of 100 bp  
 \* 41760 42434: contig of 675 bp in length  
 \* 42435 42534: gap of 100 bp  
 \* 42535 43207: contig of 673 bp in length  
 \* 43208 43307: gap of 100 bp  
 \* 43308 43980: contig of 673 bp in length  
 \* 43981 44080: gap of 100 bp  
 \* 44081 44768: contig of 688 bp in length  
 \* 44769 44868: gap of 100 bp  
 \* 44869 45556: contig of 688 bp in length  
 \* 45557 45656: gap of 100 bp  
 \* 45657 46348: contig of 692 bp in length  
 \* 46349 46448: gap of 100 bp  
 \* 46449 47143: contig of 695 bp in length  
 \* 47144 47243: gap of 100 bp  
 \* 47244 47943: contig of 700 bp in length  
 \* 47944 48043: gap of 100 bp  
 \* 48044 48727: contig of 684 bp in length  
 \* 48728 48827: gap of 100 bp  
 \* 48828 49540: contig of 713 bp in length  
 \* 49541 49640: gap of 100 bp  
 \* 49641 50332: contig of 692 bp in length  
 \* 50333 50432: gap of 100 bp  
 \* 50433 51127: contig of 695 bp in length  
 \* 51128 51227: gap of 100 bp  
 \* 51228 51915: contig of 688 bp in length  
 \* 51916 52015: gap of 100 bp  
 \* 52016 52678: contig of 663 bp in length  
 \* 52679 52778: gap of 100 bp

\* 52779 53460: contig of 682 bp in length  
 \* 53461 53560: gap of 100 bp  
 \* 53561 54251: contig of 691 bp in length  
 \* 54252 54351: gap of 100 bp  
 \* 54352 54921: contig of 570 bp in length  
 \* 54922 55021: gap of 100 bp  
 \* 55022 55702: contig of 681 bp in length  
 \* 55703 55802: gap of 100 bp

Query Match 92.4%; Score 19.4; DB 2; Length 149710;  
 Best Local Similarity 95.2%; Pred. No. 4.7;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atagggtggaggctatttgg 21  
 | ||||| ||||| ||||| |||||  
 Db 47881 ACAGGTTGGGAGGCTATTGG 47861

RESULT 13  
 AC019254/c  
 LOCUS  
 DEFINITION  
 AC019254  
 ACCESSION  
 AC019254.6 GI:15148226  
 KEYWORDS  
 HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 179183)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 15, clone RP11-424I19  
 Unpublished  
 2 (bases 1 to 179183)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bede,F.,  
 Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G., Castle,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 Dearrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,  
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 11, 2001 this sequence version replaced gi:14029977.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L2454  
 Center clone name: 424\_I\_19

TITLE  
 JOURNAL  
 COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.





```

* 146081 163677: contig of 17597 bp in length
* 163678 163777: gap of 100 bp
* 163778 167409: contig of 3632 bp in length.
FEATURES
  Source
    1..167409
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="2"
      /map="2"
      /clone="RP11-44108"
      /clone_lib="RPC1-11 Human Male BAC"
      1..11391
        /note="assembly_fragment
        clone_end:SP6
        vector_side:left"
        11492..13778
          /note="assembly_fragment"
        13879..15678
          /note="assembly_fragment"
        15779..17278
          /note="assembly_fragment"
        17379..19365
          /note="assembly_fragment"
        19466..22846
          /note="assembly_fragment"
        22947..26400
          /note="assembly_fragment"
        26501..28910
          /note="assembly_fragment"
        29011..51360
          /note="assembly_fragment"
        51461..55176
          /note="assembly_fragment"
        55277..58304
          /note="assembly_fragment"
        58405..62806
          /note="assembly_fragment"
        62907..69285
          /note="assembly_fragment"
        69386..79212
          /note="assembly_fragment"
        79313..87095
          /note="assembly_fragment"
        87196..95331
          /note="assembly_fragment"
        95432..102696
          /note="assembly_fragment"
        102797..111775
          /note="assembly_fragment"
        111876..122023
          /note="assembly_fragment"
        122124..132655
          /note="assembly_fragment"
        132756..145980
          /note="assembly_fragment"
        146081..163677
          /note="assembly_fragment"
        163778..167409
          /note="assembly_fragment
          clone_end:T7
          vector_side:right"

```

```

RESULT 15
AC013820
LOCUS
DEFINITION
AC013820
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

AC013820 179510 bp DNA HTG 01-APR-2000
Homo sapiens clone RP11-21P24, WORKING DRAFT SEQUENCE, 8 unordered
pieces.
AC013820
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 179510)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-21P24
Unpublished
2 (bases 1 to 179510)

```

```

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo.A., Wheeler,J., Wu,X.,
Wyman,B., Ye.W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-NOV-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 1, 2000 this sequence version replaced gi:6573887.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4081
Center clone name: 21_P_24
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 175240 bases at least Q40
Consensus quality: 176804 bases at least Q30
Consensus quality: 177720 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 178810; sum-of-contigs
Quality coverage: 6.8 in Q20 bases; agarose-fp
Quality coverage: 7.0 in Q20 bases; sum-of-contigs

```

```

TITLE
JOURNAL
COMMENT

```

```

NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2601: contig of 2601 bp in length
* 2602 2701: gap of 100 bp
* 2702 8297: contig of 5596 bp in length
* 8298 8397: gap of 100 bp
* 8398 27253: contig of 18856 bp in length
* 27254 27353: gap of 100 bp

```

```

BASE COUNT 51934 a 30270 c 30508 g 52497 t 2200 others
ORIGIN

```

```

Query Match 84.8%; Score 17.8; DB 2; Length 167409;
Best Local Similarity 90.5%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 atagggtggaggctatttgg 21
||||| ||||| ||||| |||||
Db 126062 ATAGGTTGGAGGCTATTTCG 126042

```

\* 27354 41791: contig of 14438 bp in length  
\* 41792 41891: gap of 100 bp  
\* 41892 61037: contig of 19146 bp in length  
\* 61038 61137: gap of 100 bp  
\* 61138 88113: contig of 26976 bp in length  
\* 88114 88213: gap of 100 bp  
\* 88214 131619: contig of 43406 bp in length  
\* 131620 131719: gap of 100 bp  
\* 131720 179510: contig of 47791 bp in length.

FEATURES

source  
1. .179510  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-21P24"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .2601  
/note="assembly\_fragment"  
misc\_feature 2702. .8297  
/note="assembly\_fragment"  
misc\_feature 8398. .27253  
/note="assembly\_fragment"  
misc\_feature 27354. .41791  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right"  
41892. .61037  
/note="assembly\_fragment"  
misc\_feature 61138. .88113  
/note="assembly\_fragment"  
misc\_feature 88214. .131619  
/note="assembly\_fragment"  
clone\_end:r7  
vector\_side:right"  
131720. .179510  
/note="assembly\_fragment"  
BASE COUNT 59127 a 32374 c 31805 g 55501 t 703 others  
ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 179510;  
Best Local Similarity 90.5%; Pred. No. 39;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 atagggtggaggctattgg 21  
|||||  
Db 107231 ATAGAGTTGGAGGCTATTGG 107251  
  
Search completed: February 15, 2002, 18:46:57  
Job time: 20056 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:01:09 ; Search time 868.33 Seconds  
(without alignments)  
20.734 Million cell updates/sec

Title: US-09-698-903b-3

Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	21	100.0	21	22	PCR primer MLD008
2	21	100.0	3265	10	TA29 gene insert i
3	21	100.0	4832	22	Nucleotide sequenc
4	21	100.0	4946	18	T-DNA of plasmid p
5	21	100.0	4946	22	Nucleotide sequenc
6	21	100.0	5864	17	Plasmid pTCOLL3 T-
7	21	100.0	5865	22	Chimeric T-DNA of
8	16.8	80.0	1983	22	Dengue-2 virus der
9	16.8	80.0	10723	18	CDNA encoding poly
10	16.8	80.0	10723	18	CDNA sequence enco
11	16.4	78.1	694	22	Human lung tumour

c 12	16.4	78.1	857	14	AAQ46850	Recombinant human
c 13	16.4	78.1	857	17	AAAT14564	Human milk kappa-c
c 14	16.4	78.1	858	16	AAO89598	Human kappa-caseln
c 15	16.4	78.1	2185	20	AAZ21101	Human secreted pro
c 16	16.4	78.1	13104	14	AAQ46852	Clone of recombina
c 17	16.2	77.1	1418	20	AAQ46867	rodent DCMPI C-lec
c 18	16.2	77.1	24978	20	AAQ60209	SEQ ID 3 of US5914
c 19	16.2	77.1	24979	21	AAA52321	Genomic DNA sequen
c 20	15.8	75.2	394	21	AAQ03474	Human secreted pro
c 21	15.8	75.2	660	21	AAQ12948	Aspergillus oryzae
c 22	15.8	75.2	777	15	AAQ070891	Tomato spotted wil
c 23	15.8	75.2	1026	16	AAQ94047	Human ALD gene exo
c 24	15.8	75.2	1549	22	AAH14886	Human cDNA sequenc
c 25	15.8	75.2	1556	22	AAH74185	s1 RNA binding reg
c 26	15.8	75.2	1577	22	AAH59214	Human polynucleoti
c 27	15.8	75.2	1591	22	AAH16816	Human cDNA sequenc
c 28	15.8	75.2	1602	22	AAH15727	Human cDNA sequenc
c 29	15.8	75.2	1643	19	AAV43711	Human sodium-depen
c 30	15.8	75.2	1644	21	AAQ76111	Human ORFX ORF1666
c 31	15.8	75.2	1795	19	AAV57910	Human haemochromat
c 32	15.8	75.2	3049	15	AAQ070890	Tomato spotted wil
c 33	15.8	75.2	9139	21	AAZ35274	Soybean retroeleme
c 34	15.4	73.3	2338	22	AAH16691	Human cDNA sequenc
c 35	15.4	73.3	3024	20	AAZ42247	Human normal biadd
c 36	15.4	73.3	4636	19	AAV23920	Human alpha3 integ
c 37	15.4	73.3	5769	19	AAV15004	Receptor protein t
c 38	15.4	73.3	4453	20	AAZ23519	Human kidney amino
c 39	15.2	72.4	139	12	AAQ13998	Template DNA. Syn
c 40	15.2	72.4	140	14	AAQ35327	DNA amplification
c 41	15.2	72.4	455	22	AAI12272	Probe #2205 for ge
c 42	15.2	72.4	455	22	AAI33627	Probe #2313 used t
c 43	15.2	72.4	455	22	AAI02189	Probe #2180 used t
c 44	15.2	72.4	649	21	AAFI4117	Aspergillus oryzae
c 45	15.2	72.4	1302	21	AAC50359	Arabidopsis thalia

ALIGNMENTS

RESULT 1  
AAD06992  
ID AAD06992 standard; DNA; 21 BP.  
XX AC  
XX AAD06992;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE PCR primer MLD008 to amplify a fragment of pTCOLL3.  
XX

KW Plasmid pTCOLL3; transgenic Brassica plant; transformation event;  
KW male-sterility gene; tobacco; PCR primer; probe; ss.  
XX

OS Chimeric - Bacillus amyloliquefaciens.

OS Chimeric - Nicotiana tabacum.

XX WO200131042-A2.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-EPI0680.

XX 29-OCT-1999; 99US-0430497.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX Weston B, De Beuckeleer M;

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX

PS Claim 1; Page 26; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is PCR primer MLD008 used to amplify a fragment of  
 CC plasmid pTCO113. The primer is also used as a probe. The amplified  
 CC fragment comprises coding region of barnase from  
 CC Bacillus amyloliquefaciens and promoter region of the anther-specific  
 CC gene TA29 from Nicotiana tabacum. This primer corresponds to position  
 CC 4697-4717 of plasmid pTCO113.

XX Sequence 21 BP; 4 A; 1 C; 10 G; 6 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.27; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0;

Qy 1 atagggtggaggctattgg 21  
 Db 1 atagggtggaggctattgg 21

RESULT 2

AA92188  
 ID AAN92188 standard; DNA; 3265 BP.  
 XX  
 AC AAN92188;

DT 11-APR-1990 (first entry)

DE TA29 gene insert in pTA29S3.

KW TA29 promoter; pTA29S3;

OS Nicotiana tabacum "Samsun".

Key	Location/Qualifiers
FT TATA-signal	1445..1451
FT	/*tag= a
FT CDS	1526..4991
FT CDS	/*tag= b
FT	1940..2296
FT	/*tag= c

PN WO8910396-A.

PD 02-NOV-1989.

PF 27-APR-1989; 89WO-EP00495.

PR 28-APR-1988; 88GB-0010120.

PA (PLAN-) PLANT GENETIC SYSTEMS NV.

PI Mariani C, Leemans J, De Greef W, De Beuckeleer M;

XX WPI; 1989-339961/46.

DR P-PSDB; AAP93313.

PT Transformed plant cell with disrupted metabolism - from a product encoded  
 PT by foreign male-sterility DNA inserted in the nuclear genome.

PS Example 2; fig. 3A; 89pp; English.

XX Plasmid pTA29S3 was constructed subcloning a ClaI fragment contg. the  
 CC TA29 gene from lambda TA29 into pUK31. The TA29 promoter is used in a  
 CC vector for transforming plant cells. Tag b is the coding sequence of the

CC TA29 gene; tag b is the cDNA insert from pBR329 (AAN92187).  
 XX  
 SQ Sequence 3265 BP; 1016 A; 581 C; 623 G; 1045 T; 0 other;

Query Match 100.0%; Score 21; DB 10; Length 3265;  
 Best Local Similarity 100.0%; Pred. No. 0.41;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atagggtggaggctattgg 21  
 Db 177 atagggtggaggctattgg 197

RESULT 3

AAH25423/C  
 ID AAH25423 standard; DNA; 4832 BP.

XX  
 AC AAH25423;

DT 22-AUG-2001 (first entry)

DE Nucleotide sequence of plasmid pTHW118.

XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barstar gene; ss.  
 XX

OS Synthetic.

OS Streptomyces hygroscopicus.

OS Arabidopsis thaliana.

OS Bacillus amyloliquefaciens.

OS Nicotiana tabacum.

Key	Location/Qualifiers
FT misc_feature	1..25
FT	/*tag= a
FT	/note= "right border repeat from TL-DNA from pTIB6S3"
FT misc_feature	26..53
FT	/*tag= b
FT	/note= "synthetic polylinker derived sequences"
FT misc_feature	54..90
FT	/*tag= c
FT	/note= "residual sequence from TL-DNA at right border repeat"
FT misc_feature	91..97
FT	/*tag= d
FT	/note= "synthetic polylinker derived sequences"
FT 3'UTR	complement (98..309)
FT	/*tag= e
FT misc_feature	310..330
FT	/note= "3' UTR from TL-DNA gene 7 of pTIB6S3"
FT	/*tag= f
FT	/note= "synthetic polylinker derived sequences"
FT CDS	complement (331..882)
FT	/*tag= g
FT	/note= "Streptomyces hygroscopicus bialaphos resistance (bar) gene"
FT promoter	complement (883..2608)
FT	/*tag= h
FT	/note= "atS1A ribulose-1,5-biphosphate carboxylase small subunit gene from Arabidopsis thaliana"
FT misc_feature	2609..2658
FT	/*tag= i
FT	/note= "synthetic polylinker derived sequences"
FT 3'UTR	complement (2659..2919)
FT	/*tag= j
FT	/note= "taql fragment from 3' UTR of nopaline synthase gene from T-DNA of pTIT37 and containing plant polyadenylation signals"
FT misc_feature	2920..2940
FT	/*tag= k
FT	/note= "synthetic polylinker derived sequences"
FT 3'UTR	2941..2980

```

FT FT /*tag= l
FT FT /note= "downstream of Bacillus amyloliquefaciens
FT FT barstar coding region"
FT FT complement (2981..3253)
FT FT /*tag= m
FT FT /note= "Barstar gene coding region from Bacillus
FT FT amyloliquefaciens"
FT FT complement (3254..4762)
FT FT /*tag= n
FT FT /note= "anther-specific gene TA29 promoter from
FT FT Nicotiana tabacum"
FT FT 4763..4807
FT FT /*tag= o
FT FT /note= "synthetic polylinker derived sequences"
FT FT 4808..4832
FT FT /*tag= p
FT FT /note= "left border repeat from TL-DNA from pTiB653"
XX XX
XX PN WO200141558-A1.
XX XX
XX PD 14-JUN-2001.
XX XX
XX PF 06-DEC-2000; 2000WO-EPI2872.
XX XX
XX PR 08-DEC-1999; 99US-0457037.
XX XX
XX PA (AVET ) AVENTIS CROPS SCIENCE NV.
XX XX
XX PI De Both G, De Beuckeleer M;
XX XX
XX DR WPI; 2001-381419/40.
XX XX
XX PT Transgenic winter oilseed rape plants suited for producing hybrid seed
XX PT with improved qualities, comprises a male-sterility gene and fertility
XX PT restorer gene, integrated into the genome
XX XX
XX PS Example 1; Page 80-82; 98pp; English.
XX XX
XX CC The specification describes a pair of transgenic winter oilseed rape
XX CC plants suited for producing hybrid seed. One of the plants has an
XX CC expression cassette comprising a male-sterility gene, and the other
XX CC plant has an expression cassette comprising a fertility restorer gene,
XX CC integrated into the genome. The fertility restorer gene is capable of
XX CC preventing the activity of the male-sterility gene. The plant pair is
XX CC useful for producing hybrid seed. Plants developed from the hybrid
XX CC seed have agronomic performance, genetic stability and adaptability to
XX CC different genetic backgrounds. The present sequence represents
XX CC plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
XX CC a fertility restorer gene. The plasmid is used to create transgenic
XX CC plants of the invention.
XX SQ Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;

Query Match 100.0%; Score 21; DB 22; Length 4832;
Best Local Similarity 100.0%; Pred. NO. 0.42;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttg 21
|||||
Db 4602 ATAGGGTGGGAGGCTATTGG 4582

RESULT 4
AAT59531/C
ID AAT59531 standard; DNA; 4946 BP.
XX AC AAT59531;
XX XX
XX DT 07-MAY-1997 (first entry)
XX XX T-DNA of plasmid pTHW107.
XX DE
XX XX

```

```

KW KW Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PARP;
KW KW niacinamide; Agrobacterium; T-DNA; male sterile; barnase;
KW KW ribonuclease; RNase; cereal; wheat; oilseed rape; Brassica napus;
XX XX plasmid pTHW107; ds.
OS OS Chimeric Agrobacterium sp.;
OS OS Chimeric Arabidopsis thaliana;
OS OS Chimeric Nicotiana tabacum.
XX XX
XX FT Key Location/Qualifiers
XX FT misc_RNA complement (1..25)
XX FT /*tag= a
XX FT /label= RB
XX FT /note= "T-DNA right border"
XX FT complement (97..330)
XX FT /*tag= b
XX FT /label= 3'g7
XX FT /note= "3' untranslated region contg. the poly-A
XX FT complement (331..882)
XX FT CDS signal of gene-7 of Agrobacterium T-DNA"
XX FT complement (331..882)
XX FT /*tag= c
XX FT /label= Bar
XX FT /product= phosphinothricin acetyltransferase
XX FT complement (883..2608)
XX FT /*tag= d
XX FT /label= PSSU
XX FT /note= "promoter region of Rubisco small subunit
XX FT gene of Arabidopsis thaliana"
XX FT complement (2658..3031)
XX FT /*tag= e
XX FT /label= 3'nos
XX FT /note= "3' untranslated region contg. the poly-A
XX FT signal of the nopaline-synthase gene of
XX FT Agrobacterium T-DNA"
XX FT complement (3032..3367)
XX FT /*tag= f
XX FT /label= Barnase
XX FT /product= barnase
XX FT complement (3368..4876)
XX FT /*tag= g
XX FT /label= pTA29
XX FT /note= "promoter region of tobacco pTA29 gene"
XX FT complement (4822..4946)
XX FT /*tag= h
XX FT /label= LB
XX FT /note= "T-DNA left border"
XX XX
XX PN EP757102-A1.
XX XX
XX PD 05-FEB-1997.
XX XX
XX PF 04-AUG-1995; 95EP-0401844.
XX XX
XX PR 04-AUG-1995; 95EP-0401844.
XX XX
XX PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX XX
XX PI De Block M;
XX XX
XX DR WPI; 1997-111050/11.
XX XX
XX PT Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase
XX PT inhibitor - reduces the cultured cells response to stress and
XX PT reduces metabolism
XX PS Example 3; Page 13-16; 25pp; English.
XX XX
XX CC Plasmid pTHW107 is a vector carrying T-DNA (AAT59531) comprising a
XX CC barnase coding sequence under control of the tobacco TA29 gene
XX CC stamen-specific promoter and a phosphinothricin acetyltransferase
XX CC coding sequence under control of an Arabidopsis Rubisco small
XX CC subunit gene promoter. Oilseed rape hypocotyl explants were
XX CC infected with Agrobacterium tumefaciens C58C1Rif carrying vector

```

CC pTHW107 and helper T1 plasmid pMP60. In some treatments, the  
 CC hypocotyls were treated with the poly-(ADP-ribose) polymerase  
 CC inhibitor niacinamide (250 mg/l) 4 days prior to infection.  
 CC plants regenerated from niacinamide-treated transformed calli  
 CC had a low copy number and displayed less variation in the  
 CC expression profile of the transgenes.  
 SQ Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;  
 Query Match 100.0%; Score 21; DB 18; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.42;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 atagggtggaggctattgg 21  
 |||||  
 Db 4716 ATAGGTGGGAGGCTATTGG 4696  
 RESULT 5  
 AAH25422/C  
 ID AAH25422 standard; DNA; 4946 BP.  
 AC AAH25422;  
 XX 22-AUG-2001 (first entry)  
 DT Nucleotide sequence of plasmid pTHW107.  
 DE Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 XX fertility restorer gene; barnase gene; ss.  
 KW Synthetic.  
 XX Streptomyces hygroscopicus.  
 OS Arabidopsis thaliana.  
 OS Bacillus amyloliquefaciens.  
 OS Nicotiana tabacum.  
 XX Key Location/Qualifiers  
 FT misc\_feature 1..25  
 FT /tag= a  
 FT /note= "right border repeat from TL-DNA from pTIB6S3"  
 FT misc\_feature 26..97  
 FT /tag= b  
 FT /note= "synthetic polylinker derived sequences"  
 FT 3'UTR complement (98..309)  
 FT /tag= c  
 FT /note= "3' UTR from TL-DNA gene 7 of pTIB6S3"  
 FT misc\_feature 310..330  
 FT /tag= d  
 FT /note= "synthetic polylinker derived sequences"  
 FT CDS 331..882  
 FT /tag= e  
 FT /note= "Streptomyces hygroscopicus bar gene"  
 FT promoter complement (883..2608)  
 FT /tag= f  
 FT /note= "atsIA ribulose-1,5-biphosphate carboxylase  
 FT small subunit gene from Arabidopsis thaliana"  
 FT misc\_feature 2609..2658  
 FT /tag= g  
 FT /note= "synthetic polylinker derived sequences"  
 FT 3'UTR complement (2659..2919)  
 FT /tag= h  
 FT /note= "Taql fragment from 3' UTR of nopaline  
 FT synthase gene from T-DNA of pTIT37 and  
 FT containing plant polyadenylation signals"  
 FT 3'UTR 2920..3031  
 FT /tag= i  
 FT /note= "3'UTR downstream of Bacillus amyloliquefaciens  
 FT barnase coding region"  
 FT CDS complement (3032..3367)  
 FT /tag= j  
 FT /note= "Barnase coding region from Bacillus

FT amyloliquefaciens"  
 FT complement (3368..4877)  
 FT /\*tag= k  
 FT /note= "anther-specific gene TA29 promoter from  
 FT Nicotiana tabacum"  
 FT misc\_feature 4878..4921  
 FT /tag= l  
 FT /note= "synthetic polylinker derived sequences"  
 FT misc\_feature 4922..4946  
 FT /tag= m  
 FT /note= "left border repeat from TL-DNA from pTIB6S3"  
 XX WO200141558-A1.  
 PN 14-JUN-2001.  
 XX 06-DEC-2000; 2000WO-EP12872.  
 XX 08-DEC-1999; 99US-0457037.  
 PR (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX De Both G, De Beuckeleer M;  
 PI WPI; 2001-381419/40.  
 DR Transgenic winter oilseed rape plants suited for producing hybrid seed  
 XX with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome -  
 XX Example 1; Page 78-80; 98pp; English.  
 PS The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene, and the other  
 CC plant has an expression cassette comprising a fertility restorer gene,  
 CC integrated into the genome. The fertility restorer gene is capable of  
 CC preventing the activity of the male-sterility gene. The plant pair is  
 CC useful for producing hybrid seed. Plants developed from the hybrid  
 CC seed have agronomic performance, genetic stability and adaptability to  
 CC different genetic backgrounds. The present sequence represents  
 CC plasmid pTHW107. This plasmid comprises the barnase gene, which acts as  
 CC a male-sterility gene. The plasmid is used to create transgenic plants  
 XX of the invention.  
 SQ Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;  
 Query Match 100.0%; Score 21; DB 22; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.42;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 atagggtggaggctattgg 21  
 |||||  
 Db 4716 ATAGGTGGGAGGCTATTGG 4696  
 RESULT 6  
 AAT39339/C  
 ID AAT39339 standard; DNA; 5864 BP.  
 XX AAT39339;  
 AC AAT39339;  
 XX 22-JAN-1997 (first entry)  
 DT Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.  
 DE Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;  
 XX transgenic plant; oilseed rape; canole; Brassica napus; ds.  
 KW Synthetic.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FH



```

FT misc_feature complement (1..25)
FT /tag= a
FT /label= RB
FT /note= "right border of Agrobacterium T-DNA"
FT polyA_signal complement (98...330)
FT /tag= b
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT CDS complement (331..882)
FT /tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin
FT acetyltransferase"
FT promoter complement (883..2608)
FT /tag= d
FT /label= Psu
FT /note= "promoter of Arabidopsis Rubisco small
FT polyA_signal complement (2659..3031)
FT /tag= e
FT /label= 3'nos
FT /note= "region containing polyA signal of nopaline
FT synthase gene of Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /tag= f
FT /label= Barnase
FT /note= "Bacillus amyloliquefaciens barnase coding
FT promoter complement (3368..4877)
FT /tag= g
FT /label= PTA29
FT /note= "promoter of stamen-specific TA29 gene of
FT Nicotiana tabacum"
FT promoter 4924..5216
FT /tag= h
FT /label= Pnos
FT /note= "promoter of nopaline synthase gene of
FT Agrobacterium T-DNA"
FT CDS 5217..5489
FT /tag= i
FT /label= Barstar
FT /note= "region coding for barstar of Bacillus
FT amyloliquefaciens"
FT polyA_signal 5490..5765
FT /tag= j
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT misc_feature complement (5840..5864)
FT /tag= k
FT /label= LB
FT /note= "left border of Agrobacterium T-DNA"
FT
FT
FT WO9626283-A1.
FT
FT 29-AUG-1996.
FT
FT 21-FEB-1996; 96WO-EP00722.
FT
FT 21-FEB-1995; 95EP-0400364.
FT
FT (PLBZ ) PLANT GENETIC SYSTEMS NV.
FT
FT Botterman J, Cornelissen M, Michiels F;
FT WPI; 1996-402373/40.
FT
FT Prodn. of male sterile plants by transforming with a chimaeric
FT construct - comprising a male sterility DNA e.g. barnase and a
FT co-regulating gene, e.g. barstar, into the nuclear genome, useful
FT for generating hybrid cultivars
FT

```

```

PS Example 3; Page 33-3743-47; 56pp; English.
XX
CC Plasmid pTColl13 (AAT39339) is a T-DNA vector containing a bar gene
CC under control of the PSU promoter, a barnase gene under control
CC of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
CC gene under control of the Pnos promoter. 87% Of oilseed rape
CC plants regenerated after Agrobacterium-mediated transformation
CC using pTColl13 were male sterile. Barnase expression disturbed the
CC function of stamen cells leading to male sterility. Constitutive
CC expression of barstar counteracted any low level expression of
CC barnase in non-stamen tissue.
XX
SQ Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;

Query Match 100.0%; Score 21; DB 17; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttg 21
Db 4716 ATAGGGTGGGAGGCTATTGG 4696
|||||
RESULT 7
AAD06990/c
ID AAD06990 standard; DNA; 5865 BP.
XX
AC AAD06990;
XX
XX 06-AUG-2001 (first entry)
XX
XX Chimeric T-DNA of plasmid pTColl13.
XX
XX T-DNA; plasmid pTColl13; transgenic Brassica plant; transformation event;
XX male-sterility gene; chimeric; tobacco; ds.
XX
XX Chimeric - Streptomyces hygroscopicus.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Bacillus amyloliquefaciens.
OS Chimeric - Nicotiana tabacum.
OS Chimeric - Agrobacterium tumefaciens.
OS Chimeric - Unidentified.
XX
XX Key Location/Qualifiers
FT misc_feature 1..25 /tag= a
FT /note= "Right border repeat from the TL-DNA from
FT pTiB6S3"
FT misc_feature 26..53 /tag= b
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 54..90 /tag= c
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 3'UTR 98..309 /tag= d
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT misc_feature 310..331 /tag= e
FT /note= "Synthetic polylinker derived sequence".
FT CDS 332..883
FT /tag= f
FT /product= "Protein encoded by bialaphos resistance
FT gene (bar) of Streptomyces hygroscopicus"
FT promoter 884..2609
FT /tag= g
FT /note= "Promoter from the atsIA ribulose-1,5-bisphosphate
FT carboxylase small subunit gene from Arabidopsis thaliana"
FT misc_feature 2610..2659
FT /tag= h
FT

```

FT misc\_feature /note= "Synthetic polylinker derived sequence"  
 FT 2660..2920  
 FT /\*tag= i  
 FT /note= "TagI fragment from the 3' untranslated end of the  
 FT nopaline synthase gene (3'nos) from the T-DNA of pTi37  
 FT and containing plant polyadenylation signals"  
 FT 2921..2936  
 FT /\*tag= j  
 FT /note= "Synthetic polylinker derived sequence"  
 FT 2937..3032  
 FT /\*tag= k  
 FT /note= "The 3' untranslated region downstream from the  
 FT barnase coding sequence of Bacillus amyloliquefaciens"  
 FT 3033..3368  
 FT /\*tag= l  
 FT /product= "Protein encoded by barnase gene from  
 FT Bacillus amyloliquefaciens"  
 FT 3369..4878  
 FT /\*tag= m  
 FT /note= "Promoter region of the anther-specific gene  
 FT TA29 from Nicotiana tabacum"  
 FT 4879..4924  
 FT /\*tag= n  
 FT /note= "Synthetic polylinker derived sequence"  
 FT 4925..5215  
 FT /\*tag= o  
 FT /note= "Promoter of the nopaline synthase gene from the  
 FT T-DNA of pTi37 of Agrobacterium tumefaciens"  
 FT 5216..5217  
 FT /\*tag= p  
 FT /note= "Synthetic polylinker derived sequence"  
 FT 5218..5490  
 FT /\*tag= q  
 FT /product= "Protein encoded by barstar gene of  
 FT Bacillus amyloliquefaciens"  
 FT 5491..5530  
 FT /\*tag= r  
 FT /note= "Sequence from the 3' untranslated end of the  
 FT barstar gene from Bacillus amyloliquefaciens"  
 FT 5531..5554  
 FT /\*tag= s  
 FT /note= "Synthetic polylinker derived sequence"  
 FT 5555..5766  
 FT /\*tag= t  
 FT /note= "The 3' untranslated end from the TL-DNA  
 FT gene 7 (3'g7) of pTiB6S3"  
 FT 5767..5773  
 FT /\*tag= u  
 FT /note= "Synthetic polylinker derived sequence"  
 FT 5774..5810  
 FT /\*tag= v  
 FT /note= "Residual sequence from the TL-DNA at the  
 FT right border repeat"  
 FT 5811..5840  
 FT /\*tag= w  
 FT /note= "Synthetic polylinker derived sequence"  
 FT 5841..5865  
 FT /\*tag= x  
 FT /note= "Left border repeat from the TL-DNA from  
 FT pTiB6S3"  
 FT  
 PN WO200131042-A2.  
 XX  
 PD 03-MAY-2001.  
 XX  
 XX 26-OCT-2000; 2000WO-EP10680.  
 PF  
 XX 29-OCT-1999; 99US-0430497.  
 PR  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA  
 XX Weston B, De Beuckeleer M;  
 PI  
 XX

DR WPI; 2001-300517/31.  
 XX  
 FT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX  
 PS Claim 1; Page 47-49; 53pp; English.  
 XX  
 CC The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is chimeric T-DNA of plasmid pTCol13. This sequence  
 CC comprises right border repeat, left border repeat and 3' untranslated  
 CC region (UTR) from TL-DNA of pTiB6S3, synthetic polylinker sequences,  
 CC Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens  
 CC and barstar gene from Bacillus amyloliquefaciens and promoters of atSia  
 CC ribulose-1,5-bisphosphate carboxylase small subunit gene from  
 CC Arabidopsis thaliana, the anther-specific gene TA29 from  
 CC Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTi37  
 CC of Agrobacterium tumefaciens.  
 XX  
 SQ Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;  
 Query Match 100.0%; Score 21; DB 22; Length 5865;  
 Best Local Similarity 100.0%; Pred. No. 0.43; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0;  
 QY 1 atagggtggaggtattgg 21  
 Db 4717 ATAGGGTGGAGGCTATTGG 4697  
 RESULT 8  
 AAF83820/c  
 ID AAF83820 standard; DNA; 1983 BP.  
 XX  
 AC AAF83820;  
 XX  
 DT 23-JUL-2001 (first entry)  
 XX  
 DE Dengue-2 virus derived DNA fragment.  
 XX  
 KW Yellow fever virus; prM-E protein; flavivirus; chimeric; medicament;  
 KW infection; tumor antigen; cytokine; lymphoid; reticuloendothelial;  
 KW cancer; virucide; vaccine; dengue virus; ds.  
 XX  
 OS Dengue virus.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 1..1983  
 FT /\*tag= a  
 FT  
 PN WO200139802-A1.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 01-DEC-2000; 2000WO-US32821.  
 XX  
 PR 01-DEC-1999; 99US-0452638.  
 XX  
 XX (ORAV-) ORAVAX INC.  
 PA  
 XX Chambers TJ, Monath TP, Guirakhoo F;  
 PI  
 XX WPI; 2001-343953/36.  
 DR  
 DR P-PSDB; AAB84901.  
 XX

PT Chimeric live, infectious, attenuated yellow fever viruses used for  
 PT preventing and treating diseases caused by flaviviruses have prM-E  
 PT nucleotide sequence from a second, different flavivirus as functional  
 PT yellow fever prM-E is not expressed -  
 XX  
 PS Disclosure; Page 196-198; 232pp; English.  
 XX  
 CC The invention relates to a chimeric live, infectious, attenuated virus  
 CC comprising a yellow fever virus with the nucleotide sequence encoding a  
 CC prM-E protein deleted, truncated or mutated so that functional yellow  
 CC fever virus prM-E protein is not expressed and also integrated into the  
 CC genome of the yellow fever virus a nucleotide sequence encoding a prM-E  
 CC protein of a second, different flavivirus so that the prM-E protein of  
 CC the second flavivirus is expressed. The chimeric live, infectious,  
 CC attenuated virus is used to prepare medicaments for preventing or  
 CC treating flavivirus infection in a patient. The yellow fever virus vector  
 CC produces its gene product (tumor antigen or cytokine) in cells of the  
 CC lymphoid or reticuloendothelial system or in a precursor of these systems  
 CC in patients with cancer. Flaviviruses replicate in the cytoplasm of cells  
 CC so that the virus replication does not involve integration of the viral  
 CC genome into the host cell. The present sequence represents a DNA fragment  
 CC derived from dengue-2 virus.  
 XX  
 SQ Sequence 1983 BP; 645 A; 409 C; 509 G; 420 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 1983;  
 Best Local Similarity 90.0%; Pred. No. 42;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 taggggtggaggctatttgg 21  
 ||||| ||||| |||||  
 Db 666 TAGGGTGCAGCGCTTTGG 647

RESULT 9  
 AAT49304/c  
 ID AAT49304 standard; cdna; 10723 BP.  
 XX  
 AC AAT49304;  
 XX  
 DT 12-SEP-1997 (first entry)  
 XX  
 DE cDNA encoding polyprotein of attenuated DEN-2 virus, PDK-53.  
 XX  
 KW Dengue 2 virus; polyprotein; capsid; prM; M; E; NS1; NS2A; NS2B; NS3;  
 KW NS4A; NS4B; NS5; PDK-53; quadravalent vaccine; immunity; serotype;  
 KW chimeric DEN-2/1 virus; chimeric DEN-2/3 virus; chimeric DEN-2/4 virus;  
 KW dengue fever; fatal dengue haemorrhagic fever; dengue shock syndrome;  
 KW DHF; DSS; ss.  
 XX  
 OS Dengue 2 virus, strain 16681.  
 OS Synthetic.

Key Location/Qualifiers  
 FT CDS 97..10272  
 FT /tag= a  
 FT /product= DEN-2 attenuated polyprotein  
 FT /transl\_except(pos:643..645, aa:Xaa)  
 FT /transl\_except(pos:1135..1137, aa:Xaa)  
 FT /transl\_except(pos:1393..1395, aa:Xaa)  
 FT /transl\_except(pos:2809..2811, aa:Xaa)  
 FT /transl\_except(pos:3040..3042, aa:Xaa)  
 FT /transl\_except(pos:9208..9210, aa:Xaa)  
 FT /note= "Xaa = unknown amino acid"  
 FT 57  
 FT mutation  
 FT /tag= b  
 FT /note= "C>T mutation"  
 FT 524  
 FT mutation  
 FT /tag= c  
 FT /note= "A>T mutation, causes Asp to Val substitution"  
 FT 2055  
 FT mutation  
 FT /tag= d

FT /note= "C>T mutation"  
 FT 2579  
 FT /tag= e  
 FT /note= "G>A mutation, causes Gly to Asp substitution"  
 FT 4018  
 FT /tag= f  
 FT /note= "C>T mutation, causes Leu to Phe substitution"  
 FT 5547  
 FT /tag= g  
 FT /note= "C>T mutation"  
 FT 6599  
 FT /tag= h  
 FT /note= "G>C mutation, causes Gly to Ala substitution"  
 FT 8571  
 FT /tag= i  
 FT /note= "C>T mutation"  
 XX  
 PN W09640933-A1.  
 XX  
 XX 19-DEC-1996.  
 PD  
 XX 06-JUN-1996; 96WO-US09209.  
 PF  
 XX 07-JUN-1995; 95US-0483292.  
 PR  
 XX (UYMA-) UNIV MAHIDOL AT SALAYA.  
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
 PA  
 XX Bhamarapavati N, Butrapet S, Chang J, Gubler DJ;  
 PI Halstead SB, Kinney R, Trent DW;  
 PI  
 XX WPI: 1997-052330/05.  
 DR  
 DR P-PSDB; AAW06591.  
 XX  
 PT PDK-53, a clone of infectious attenuated Dengue 2 virus strain 16681  
 PT - also chimeric DEN-2/1, DEN-2/3 and DEN-1/4 viruses, used as a  
 PT quadravalent vaccine for protecting against Dengue virus infection  
 XX  
 PS Claim 27; Page 122-136; 261pp; English.  
 XX  
 CC This sequence encodes the polyprotein from an attenuated derivative  
 CC of Dengue 2 virus, strain 16681. The derivative is designated PDK-53.  
 CC The polyprotein comprises the capsid, prM, M, E, NS1, NS2A, NS2B, NS3,  
 CC NS4A, NS4B and NS5 proteins. The PDK-53 viral sequence may be used in  
 CC the production of a quadravalent vaccine which provides immunity against  
 CC all four serotypes of dengue virus. The vaccine also comprises a  
 CC chimeric DEN-2/1 virus, a chimeric DEN-2/3 virus, and/or a chimeric  
 CC DEN-2/4 virus. The new quadravalent vaccines are used to protect  
 CC against infection by all four serotypes of dengue virus, DEN-1, DEN-2,  
 CC DEN-3 and DEN-4, which can lead to dengue fever or fatal dengue  
 CC haemorrhagic fever/dengue shock syndrome (DHF/DSS). Host cells are  
 CC used to produce the recombinant protein products of the DNA constructs  
 CC which are used in the vaccines.  
 XX  
 SQ Sequence 10723 BP; 3548 A; 2196 C; 2711 G; 2261 T; 7 other;

Query Match 80.0%; Score 16.8; DB 18; Length 10723;  
 Best Local Similarity 90.0%; Pred. No. 48;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 2 taggggtggaggctatttgg 21  
 ||||| ||||| |||||  
 Db 1104 TAGGGTGCAGCGCTTTGG 1085  
 RESULT 10  
 AAT49303/c  
 ID AAT49303 standard; cdna; 10723 BP.  
 XX  
 AC AAT49303;  
 XX  
 DT 11-SEP-1997 (first entry)

```

XX DE CDNA sequence encoding polyprotein of DEN-2 virus, strain 16681.
XX KW Dengue 2 virus; polyprotein; capsid; prM; M; E; NS1; NS2A; NS2B; NS3;
XX KW NS4A; NS4B; NS5; PDK-53; quadravalent vaccine; immunity; serotype;
XX KW chimeric DEN-2/1 virus; chimeric DEN-2/3 virus; chimeric DEN-2/4 virus;
XX KW dengue fever; fatal dengue haemorrhagic fever; dengue shock syndrome;
XX KW DHF; DSS; ss.
XX OS Dengue 2 virus, strain 16681.
XX FH Key Location/Qualifiers
XX FT 97..10272
XX FT /*tag= a
XX FT /product= DEN-2 polyprotein
XX FT /trans_except(pos:9208..9210, aa:Xaa)
XX FT /note= "Xaa = unknown amino acid"
XX PN WO9604933-A1.
XX XX
XX PD 19-DEC-1996.
XX PF
XX PP 06-JUN-1996; 96WO-US09209.
XX PR
XX PR 07-JUN-1995; 95US-0483292.
XX PR (UYMA-) UNIV MAHIDOL AT SALAYA.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Bhamarapravati N, Butrapet S, Chang J, Gubler DJ;
XX PI Halstead SB, Kinney R, Trent DW;
XX XX
XX DR WPI; 1997-052330/05.
XX DR P-PSDB; AAW06590.
XX XX
XX PT PDK-53, a clone of infectious attenuated Dengue 2 virus strain 16681
XX PT - also chimeric DEN-2/1, DEN-2/3 and DEN-1/4 viruses, used as a
XX PT quadravalent vaccine for protecting against Dengue virus infection
XX PS Claim 23; Page 107-121; 261pp; English.
XX CC This sequence encodes the polyprotein from Dengue 2 virus, strain 16681.
XX CC The polyprotein comprises the capsid, prM, M, E, NS1, NS2A, NS2B, NS3,
XX CC NS4A, NS4B and NS5 proteins. A clone of this wildtype viral sequence,
XX CC PDK-53, may be used in the production of a quadravalent vaccine which
XX CC provides immunity against all four serotypes of dengue virus. The
XX CC vaccine also comprises a chimeric DEN-2/1 virus, a chimeric DEN-2/3
XX CC virus, and/or a chimeric DEN-2/4 virus. The new quadravalent vaccines
XX CC are used to protect against infection by all four serotypes of dengue
XX CC virus, DEN-1, DEN-2, DEN-3 and DEN-4, which can lead to dengue fever or
XX CC fatal dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS). Host
XX CC cells are used to produce the recombinant protein products of the DNA
XX CC constructs which are used in the vaccines.
XX SQ Sequence 10723 BP; 3553 A; 2200 C; 2713 G; 2255 T; 2 other;

Query Match 80.0%; Score 16.8; DB 18; Length 10723;
Best Local Similarity 90.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggtggaggcctattgg 21
|||||
Db 1104 TAGGCTGCAGCTCTTTGG 1085

RESULT 11
AAF68287
ID AAF68287 standard; CDNA; 694 BP.
XX AC AAF68287;
XX XX
XX DT 12-APR-2001 (first entry)

```

```

XX DE Human lung tumour protein related nucleotide sequence SEQ ID NO:205.
XX KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
XX KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
XX KW cytostatic; antisense inhibition; ss.
XX OS Homo sapiens.
XX FH WO200100828-A2.
XX FT 04-JAN-2001.
XX PF
XX PP 30-JUN-2000; 2000WO-US18061.
XX PR 30-JUN-1999; 99US-0346492.
XX PR 15-OCT-1999; 99US-0419356.
XX PR 17-DEC-1999; 99US-0468867.
XX PR 30-DEC-1999; 99US-0476300.
XX PR 06-MAR-2000; 2000US-0519642.
XX PR 22-MAR-2000; 2000US-0533077.
XX PR 10-APR-2000; 2000US-0546259.
XX PR 27-APR-2000; 2000US-0560406.
XX PR 05-JUN-2000; 2000US-0589184.
XX PR (CORI-) CORIXA CORP.
XX PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX PI Retter MW, Mannion J;
XX DR WPI; 2001-071488/08.
XX XX
XX PT Lung tumor-associated proteins and the nucleic acids that encode them,
XX PT useful for preventing, diagnosing and treating lung cancer -
XX PS Claim 4; Page 213; 436pp; English.
XX CC The present invention describes immunogenic portions of lung tumour-
XX CC associated proteins (I) and the nucleic acids (NAs) that encode them.
XX CC (I) have cytostatic activity and can be used in gene therapy, antisense
XX CC inhibition and in vaccines. The NAs and the lung tumour-associated
XX CC proteins they encode may be used in the prevention, treatment and
XX CC diagnosis of diseases associated with their inappropriate expression,
XX CC especially lung cancers. For example, the NAs may be administered to
XX CC treat diseases by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of the protein by expressing inactive proteins
XX CC or to supplement the patients own production of (I). Additionally, the
XX CC NAs may be used to produce the lung-tumour associated protein, according
XX CC to standard recombinant DNA methodology. Conversely, antisense NA
XX CC molecules may be administered to down regulate protein expression by
XX CC binding with the cells own genes and preventing their expression. The NA
XX CC and complementary sequences may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar NA sequences in
XX CC samples, and hence which patients may be in need of treatment for lung
XX CC cancer. The (I) may be used as antigens in the production of antibodies
XX CC and in assays to identify modulators (agonists and antagonists) of the
XX CC expression and activity of the protein. AAF68083 to AAF68878 and
XX CC AAF76848 to AAF76878 represent human lung tumour protein related
XX CC nucleotide and protein sequences which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 694 BP; 202 A; 111 C; 154 G; 227 T; 0 other;

Query Match 78.1%; Score 16.4; DB 22; Length 694;
Best Local Similarity 94.4%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aggggtggaggcctattgg 20
|||||
Db 290 aggggtggaggcctattgg 307

```

RESULT 12  
AAQ46850/c  
ID AAQ46850 standard; cDNA; 857 BP.  
XX AC AAQ46850;  
XX DT 26-JAN-1994 (first entry)  
XX DE Recombinant human kappa casein gene.  
XX KW Casein; supplement; milk; pharmaceutical; ss.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 45..593  
XX FT /\*tag= a  
XX FT /product= Human kappa casein.  
XX FT mat\_peptide 45..593  
XX FT /\*tag= b  
XX FT sig\_peptide 45..104  
XX FT /\*tag= c  
XX PN WO9315196-A.  
XX XX  
XX PD 05-AUG-1993.  
XX XX  
XX PF 25-JAN-1993; 93WO-DK000024.  
XX PR 23-JAN-1992; 92DK-0000088.  
XX XX  
XX PA (SYMB-) SYMBICOM AB.  
XX PI Bergstroem S, Hansson L, Hernell O, Stroemqvist M;  
XX PI Toernell J;  
XX XX  
XX DR WPI; 1993-258675/32.  
XX DR P-PSDB; AAR39351.  
XX XX  
XX PT DNA encoding human kappa-casein - used for obtaining recombinant  
XX PT polypeptide(s) for use as nutrient supplements, partic. in infant  
XX PT formulae  
XX XX  
XX PS Claim 14; Page 88-89; 110pp; English.  
XX CC The recombinant human kappa casein is produced in high yields by  
XX CC means of either a eukaryotic or prokaryotic expression system. It  
XX CC is used as a nutrient supplement in milk based products to provide a  
XX CC substantial improvement of the nutritional and biological value of  
XX CC the formulae, making it closer in similarity to human milk. It can  
XX CC also be used as a pharmaceutical.  
XX XX  
XX SQ Sequence 857 BP; 307 A; 220 C; 128 G; 202 T; 0 other;

Query Match 78.1%; Score 16.4; DB 14; Length 857;  
Best Local Similarity 94.4%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggggtggaggctatttgg 21  
||||| |||||||  
Db 359 GGGTGGGTGGCTATTGG 342

RESULT 13  
AAT14564/c  
ID AAT14564 standard; cDNA; 857 BP.  
XX AC AAT14564;  
XX XX  
XX DT 24-OCT-1996 (first entry)  
XX DE Human milk kappa-casein cDNA.

XX Milk protein; kappa-casein; rotavirus inhibition; gastroenteritis;  
KW infant formulation; immunodeficiency; diarrhoea; ss.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 45..593  
XX FT /\*tag= a  
XX XX  
XX PN WO9608269-A1.  
XX XX  
XX PD 21-MAR-1996.  
XX XX  
XX PF 05-MAY-1995; 95WO-US05676.  
XX XX  
XX PR 16-SEP-1994; 94US-0308883.  
XX PR 16-SEP-1994; 94US-0308882.  
XX XX  
XX PA (ABBO ) ABBOTT LAB.  
XX XX  
XX PI Baxter JH, Cummings RD, Mukerji P, Prieto PA, Seo AE;  
XX XX  
XX DR WPI; 1996-179724/18.  
XX XX  
XX PT Kappa-casein used as anti-rotaviral infection agent in nutritional  
XX PT product - to prevent, retard or treat rotavirus infection,  
XX PT especially in infants, children and immuno-deficient patients  
XX XX  
XX PS Claim 1; Fig 8; 42pp; English.  
XX XX  
XX CC AAT14564 encodes human milk kappa-casein which is useful as a component  
XX CC of an enteral nutritional product. Kappa-casein has anti-rotaviral  
XX CC activity, it inhibits the attachment of human rotavirus to mammalian  
XX CC cells. The kappa-casein is useful in the treatment and prevention of  
XX CC onset of gastroenteritis and other diarrhoeal diseases caused by  
XX CC rotaviridae esp. in infants, children or immunodeficient patients.  
XX CC Kappa-casein appeared to have no side effects on the subjects treated  
XX CC and is unlikely to be toxic or cause an allergic reaction.  
XX XX  
XX SQ Sequence 857 BP; 309 A; 219 C; 127 G; 202 T; 0 other;

Query Match 78.1%; Score 16.4; DB 17; Length 857;  
Best Local Similarity 94.4%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggggtggaggctatttgg 21  
||||| |||||||  
Db 359 GGGTGGGTGGCTATTGG 342

RESULT 14  
AAQ89598/c  
ID AAQ89598 standard; cDNA; 858 BP.  
XX AC AAQ89598;  
XX XX  
XX DT 06-NOV-1995 (first entry)  
XX XX  
XX DE Human kappa-casein cDNA.  
XX XX  
XX KW Kappa-casein; milk protein; ss.  
XX XX  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 107..655  
XX FT /\*tag= a  
XX XX  
XX PN US5391497-A.  
XX XX  
XX PD 21-FEB-1995.

XX 13-OCT-1992; 92US-0962569.  
 XX 13-OCT-1992; 92US-0962569.  
 XX (COLS ) UNIV COLORADO FOUND INC.  
 XX Chang Y, Ham RG, Jeffers KF, Menon RS;  
 XX WPI; 1995-160470/21.  
 XX P-PSDB; AAR72696.  
 XX DNA encoding human kappa-casein - used for the prodn. of large amts. of  
 PT highly purified kappa-casein milk protein for infant use.  
 XX Claim 4; Column 13-16; 14pp; English.  
 XX A commercial cDNA library prep'd. in lambda gtl1 from mRNA obtd. from  
 CC human breast tissue removed during the third trimester of pregnancy was  
 CC screened with rabbit anti-bovine kappa-casein cDNA. The cDNA  
 CC insert of a recombinant phage was amplified by PCR. The sequence  
 CC of an isolated full-length clone encoding human kappa-casein  
 CC (AAR72696) is given in AAQ89598.  
 XX Sequence 858 BP; 266 A; 237 C; 137 G; 218 T; 0 other;  
 SQ

Query Match 78.1%; Score 16.4; DB 16; Length 858;  
 Best Local Similarity 94.4%; Pred. No. 61;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggggtggaggctattgg 21  
 ||||| |||||  
 Db 421 GGGTGGTGGCTATTGG 404

RESULT 15  
 AAZ21101/c  
 ID AAZ21101 standard; cDNA; 2185 BP.  
 XX AAZ21101;  
 XX 17-NOV-1999 (first entry)  
 XX Human secreted protein clone qb401\_6 encoding cDNA.  
 DE Human; secreted protein; biological activity; nutritional; cytokine;  
 KW cell proliferation; differentiation; immune stimulating; vaccine;  
 KW haematopoiesis regulation; tissue growth; haemostatic; thrombolytic;  
 KW anti-inflammatory; tumour inhibition; ss.  
 XX Homo sapiens.  
 OS  
 XX WO9946287-A1.  
 PN  
 XX 16-SEP-1999.  
 PD  
 XX 11-MAR-1999; 99WO-US05243.  
 PF  
 XX 11-MAR-1998; 98US-0077521.  
 PR 14-MAY-1998; 98US-0079124.  
 PR 10-MAR-1999; 99US-0266105.  
 XX (GENY ) GENETICS INST INC.  
 PA  
 XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ;  
 XX WPI; 1999-551362/46.  
 DR P-PSDB; AAY29869.  
 DR  
 XX Polynucleotides encoding secreted human proteins, derived from human  
 PT fetal brain, human adult blood, human adult bladder, or human adult

PT neural tissue cDNA libraries. -  
 XX Claim 26; Page 109; 118pp; English.  
 XX AAZ21093 to AAZ21102 encode new human secreted proteins and AAY29861 to  
 CC AAY29873 represent the secreted proteins encoded by the polynucleotide  
 CC sequences. AAZ21103 to AAZ21112 represent probes for the secreted  
 CC proteins. The polynucleotides and proteins are predicted to have  
 CC biological activities which would make them suitable for treating,  
 CC preventing or ameliorating medical conditions in humans and animals,  
 CC although no supporting data is given. Suggested activities include  
 CC nutritional activity, cytokine and cell proliferation/differentiation  
 CC activity, immune stimulating (e.g. as vaccines) or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity,  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, and tumour  
 CC inhibition activity. The polynucleotides and proteins can also be used  
 CC as nutritional sources or supplements. Such uses include use as a protein  
 CC or amino acid supplement, use as a carbon source, use as a nitrogen  
 CC source and use as a source of carbohydrate. They may also have utility  
 CC in compositions used for bone, cartilage, tendon, ligament, and/or nerve  
 CC tissue growth or regeneration, as well as for wound healing and tissue  
 CC repair and replacement, and in the treatment of burns, incisions and  
 CC ulcers. The proteins which induce cartilage and/or bone growth in  
 CC circumstances where bone is not normally formed, have application in  
 CC the healing of bone fractures and cartilage damage or defects in humans  
 CC and other animals.  
 XX Sequence 2185 BP; 472 A; 572 C; 548 G; 593 T; 0 other;  
 SQ

Query Match 78.1%; Score 16.4; DB 20; Length 2185;  
 Best Local Similarity 94.4%; Pred. No. 66;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 taggggtggaggctattt 19  
 ||||| |||||  
 Db 1052 TAGGGTGGAGGCTATT 1035

Search completed: February 15, 2002, 19:01:13  
 Job time: 20717 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:07:18 ; Search time 353.79 Seconds  
(without alignments)  
13.443 Million cell updates/sec

Title: US-09-698-903B-3

Perfect score: 21

Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2.6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2.6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		DB ID	Description
		Match	Length		
C 1	21	100.0	4946	US-08-817-188-1	Sequence 1, Appli
C 2	21	100.0	5864	US-08-894-440-4	Sequence 4, Appli
C 3	16.4	78.1	857	US-08-308-883-1	Sequence 1, Appli
C 4	16.4	78.1	857	US-08-730-163-1	Sequence 1, Appli
C 5	16.4	78.1	857	US-08-256-799-1	Sequence 1, Appli
C 6	16.4	78.1	857	US-08-462-437-1	Sequence 1, Appli
C 7	16.4	78.1	858	US-07-962-569A-7	Sequence 7, Appli
C 8	16.4	78.1	9844	US-08-462-437-30	Sequence 30, Appl
C 9	16.4	78.1	13104	US-08-256-799-4	Sequence 4, Appli
C 10	16.4	78.1	13104	US-08-462-437-4	Sequence 4, Appli
C 11	16.2	77.1	1418	US-09-111-470-7	Sequence 7, Appli
C 12	16.2	77.1	24979	US-08-147-777-3	Sequence 3, Appli
C 13	16.2	77.1	24979	US-08-452-872-3	Sequence 3, Appli
C 14	16.2	77.1	24979	PCT-US93-03985-3	Sequence 3, Appli
C 15	15.8	75.2	1025	US-08-136-277-23	Sequence 23, Appl
C 16	15.8	75.2	1025	US-08-479-403-23	Sequence 23, Appl
C 17	15.8	75.2	1025	US-08-835-734-23	Sequence 23, Appl
C 18	15.8	75.2	1643	US-08-805-118-2	Sequence 2, Appli
C 19	15.8	75.2	1780	US-08-724-394A-19	Sequence 19, Appl
C 20	15.4	73.3	5769	US-08-652-971-1	Sequence 1, Appli
C 21	15.4	73.3	5769	US-08-991-258A-1	Sequence 1, Appli
C 22	15.4	73.3	5769	US-08-769-399-1	Sequence 1, Appli
C 23	15.4	73.3	5769	US-08-991-953A-1	Sequence 1, Appli
C 24	15.2	72.4	5538	US-08-231-193A-55	Sequence 55, Appl
C 25	15.2	72.4	5538	US-08-486-273A-55	Sequence 55, Appl
C 26	15.2	72.4	5538	US-08-940-086A-55	Sequence 55, Appl
C 27	15.2	72.4	9046	US-08-227-536-1	Sequence 1, Appli

C 28	15.2	72.4	9046	5	PCT-US95-04682-1	Sequence 1, Appli
C 29	14.8	70.5	354	2	US-08-676-279-38	Sequence 38, Appl
C 30	14.8	70.5	569	4	US-09-326-039-13	Sequence 13, Appl
C 31	14.8	70.5	2750	1	US-08-136-277-1	Sequence 1, Appli
C 32	14.8	70.5	2750	2	US-08-479-403-1	Sequence 1, Appli
C 33	14.8	70.5	2750	3	US-08-835-734-1	Sequence 1, Appli
C 34	14.6	69.5	35100	1	US-08-306-691B-19	Sequence 19, Appl
C 35	14.6	69.5	35100	5	PCT-US93-06251-19	Sequence 19, Appl
C 36	14.4	68.6	90050	4	US-09-245-041-5	Sequence 5, Appli
C 37	14.2	67.6	2899	2	US-08-624-581-2	Sequence 2, Appli
C 38	14.2	67.6	2959	3	US-08-624-581-1	Sequence 1, Appli
C 39	14.2	67.6	5994	3	US-09-032-365A-11	Sequence 11, Appl
C 40	14.2	67.6	13146	2	US-08-724-354D-3	Sequence 3, Appli
C 41	14.2	67.6	13146	3	US-09-270-984A-3	Sequence 3, Appli
C 42	14.2	67.6	14311	4	US-08-646-695-1	Sequence 1, Appli
C 43	14.2	67.6	14311	4	US-08-646-695-7	Sequence 7, Appli
C 44	14.2	67.6	14311	5	PCT-US96-06053-1	Sequence 1, Appli
C 45	14.2	67.6	14311	5	PCT-US96-06053-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-08-817-188-1/c  
; Sequence 1, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; EARLIER FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
; EARLIER FILING DATE: 1996-07-31  
; EARLIER APPLICATION NUMBER: EP 95401844.6  
; EARLIER FILING DATE: 1995-08-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; OTHER INFORMATION: plasmid pTHW107  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((1)..(25))  
; OTHER INFORMATION: T-DNA right border (RB)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((97)..(330))  
; OTHER INFORMATION: 3'gT: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
; OTHER INFORMATION: T-DNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((331)..(882))  
; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
; OTHER INFORMATION: transferase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((883)..(2608))  
; OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2658)..(3031))  
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
; OTHER INFORMATION: gene of Agrobacterium T-DNA  
; FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4876))
; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((4922)..(4946))
; OTHER INFORMATION: LB: T-DNA left border
US-08-817-188-1

Query Match      100.0%; Score 21; DB 3; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.056;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
   |||||
Db 4716 ATAGGGTGGAGGCTATTGG 4696

RESULT 2
US-08-894-440-4/c
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; FILE OF INVENTION: Method to obtain male sterile plants
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTC013
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g')
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Phos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g')
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4

Query Match      100.0%; Score 21; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
   |||||
Db 4716 ATAGGGTGGAGGCTATTGG 4696

RESULT 3
US-08-308-883-1/c
; Sequence 1, Application US/08308883
; Patent No. 5576300
; GENERAL INFORMATION:
; APPLICANT: Mukerji, P.
; APPLICANT: Prieto, P. A.
; APPLICANT: Seo, A. E.-Y.
; APPLICANT: Baxter, J. H.
; APPLICANT: Cummings, R.D.
; TITLE OF INVENTION: Method for Inhibition of Human Rotavirus Infection.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lonnie R. Drayer
; ADDRESSEE: ROSS Products Division
; ADDRESSEE: Abbott Laboratories
; STREET: 625 Cleveland Avenue
; CITY: Columbus
; STATE: Ohio
; COUNTRY: United States
; ZIP: 43215
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh System 7.1
; SOFTWARE: Clarisworks 1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,883
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: No. 5576300 applicable
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (614) 624-3774
; TELEFAX: (614) 624-3074
; TELEX: No. 5576300e
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 857 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human milk kappa-casein  
HYPOTHETICAL: NO  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: Human  
ORGANISM: Homo sapiens  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE: Adult  
HAPLOTYPE:  
TISSUE TYPE: Mammary gland  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE: Human Mammary Gland  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 45...593  
IDENTIFICATION METHOD: DNA sequencing and restriction analysis  
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the human  
PUBLICATION INFORMATION:  
AUTHORS: L. Hansson et al  
TITLE: DNA Encoding kappa-Casein, Process for Obtaining the Protein and Use The  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER: PCT/WO93/15196  
FILING DATE: 25-JAN-1993  
PUBLICATION DATE: 05-AUG-1993  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-308-883-1

Query Match 78.1%; Score 16.4; DB 1; Length 857;  
Best Local Similarity 94.4%; Pred. No. 9.5;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggggtggaggctattgg 21  
||||| |||||||  
Db 359 GGGTGGTGGCTATTGG 342

RESULT 4  
US-08-730-163-1/c  
Sequence 1, Application US/08730163  
Patent No. 5712250  
GENERAL INFORMATION:  
APPLICANT: Mukerji, P. A.  
APPLICANT: Prieto, P. A.  
APPLICANT: Seo, A. E.-Y.  
APPLICANT: Baxter, J. H.  
APPLICANT: Cummings, R. D.  
TITLE OF INVENTION: Product for Inhibition of Human Rotavirus Infection.  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lonnie R. Drayer  
ADDRESSEE: ROSS Products Division  
ADDRESSEE: Abbott Laboratories  
STREET: 625 Cleveland Avenue  
CITY: Columbus

STATE: Ohio  
COUNTRY: United States  
ZIP: 43215  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh System 7.1(D)SOFTWARE: ClarisWorks 1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/730,163  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/308,882  
FILING DATE: 16-SEP-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-3774  
TELEFAX: (614) 624-3074  
TELEX: No. 5712250e  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 857 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: Human milk kappa-casein  
HYPOTHETICAL: NO  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: Human  
ORGANISM: Homo sapiens  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE: Adult  
HAPLOTYPE:  
TISSUE TYPE: Mammary gland  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE: Human Mammary Gland  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 45...593  
IDENTIFICATION METHOD: DNA sequencing and restriction analysis  
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the h  
PUBLICATION INFORMATION:  
AUTHORS: L. Hansson et al  
TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER: PCT/WO93/15196  
FILING DATE: 25-JAN-1993  
PUBLICATION DATE: 05-AUG-1993  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-730-163-1

Query Match 78.1%; Score 16.4; DB 1; Length 857;  
Best Local Similarity 94.4%; Pred. No. 9.5;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggggtggaggctattgg 21  
||||| |||||||  
Db 359 GGGTGGTGGCTATTGG 342

RESULT 5  
US-08-256-799-1/C  
; Sequence 1, Application US/08256799  
; Patent No. 6222094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMOVIST, Mats  
; APPLICANT: BERGSTROEM, Sven  
; APPLICANT: HERNELL, Olie  
; APPLICANT: Toernell, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR  
; TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/256,799  
; FILING DATE: 08-DEC-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON=1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 45..104  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 13..44  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: 594..848  
US-08-256-799-1

Query Match 78.1%; Score 16.4; DB 4; Length 857;  
Best Local Similarity 94.4%; Pred. No. 9.5;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 995tgggaggtatttgg 21

Db 359 GGGTGGGTGGCTATTGG 342  
||||||| |||||||||  
RESULT 6  
US-08-462-437-1/C  
; Sequence 1, Application US/08462437  
; Patent No. 6232094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMOVIST, Mats  
; APPLICANT: BERGSTROEM, Sven  
; APPLICANT: HERNELL, Olie  
; APPLICANT: Toernell, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS  
; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,437  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON-1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 45..104  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 13..44  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: 594..848  
US-08-462-437-1

Query Match 78.1%; Score 16.4; DB 4; Length 857;  
Best Local Similarity 94.4%; Pred. No. 9.5;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gggtagggagctattgg 21  
||||| |||||||  
Db 359 GGGTGGGTGCTATTGG 342

RESULT 7  
US-07-962-569A-7/c  
; Sequence 7, Application US/07962569A  
; Patent No. 5391497  
; GENERAL INFORMATION:  
; APPLICANT: MENON, RAVI S.  
; APPLICANT: JEFFERS, KATHLEEN F.  
; APPLICANT: CHANG, YING-FON  
; APPLICANT: HAM, RICHARD G.  
; TITLE OF INVENTION: HUMAN K-CASEIN  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FREDERICK W. PEPPER, PH.D.  
; STREET: 11545 W. BERNARDO COURT, STE. 302  
; CITY: SAN DIEGO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92127

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/962,569A  
; FILING DATE: 19921013  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PEPPER PH.D., FREDERICK W.  
; REGISTRATION NUMBER: 31,286  
; REFERENCE/DOCKET NUMBER: 920224.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 451-1120  
; TELEFAX: (619) 451-9628  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 858 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 107..656  
US-07-962-569A-7

Query Match 78.1%; Score 16.4; DB 1; Length 858;  
Best Local Similarity 94.4%; Pred. No. 9.5;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gggtagggagctattgg 21  
||||| |||||||  
Db 421 GGGTGGGTGCTATTGG 404

RESULT 8  
US-08-462-437-30/c  
; Sequence 30, Application US/08462437  
; Patent No. 6232094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMQVIST, Mats  
; APPLICANT: BERGSTROEM, Sven  
; APPLICANT: HERNELL, Olle  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS

; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,437  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON-1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9844 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 362..425  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 426..2571  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 2572..2633  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 2634..5575  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 5576..5608  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 5609..5754  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 5755..7250  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 7251..9017  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 9018..9184  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(2580..2633, 5576..5608, 6755..7216)  
US-08-462-437-30

Query Match

78.1%; Score 16.4; DB 4; Length 9844;

```
Best Local Similarity 94.4%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gggtggggaggctatttgg 21
||||| |||||||
Db 6982 GGGTGGGTGGCTATTGG 6965

RESULT 9
US-08-256-799-4/c
; Sequence 4, Application US/08256799
; Patent No. 6222094
; GENERAL INFORMATION:
; APPLICANT: HANSSON, Lennart
; APPLICANT: STROEMQVIST, Mats
; APPLICANT: BERGSTROEM, Sven
; APPLICANT: HERNELL, Olle
; APPLICANT: Toernell, Jan
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR
; TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256,799
; FILING DATE: 06-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 88/92
; FILING DATE: 23-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: HANSSON=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: ps459
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..8834
; FEATURE:
; NAME/KEY: intron
; LOCATION: 8868..10014
; NAME/KEY: intron
; LOCATION: 10511..12277
; FEATURE:
; NAME/KEY: exon
; LOCATION: 8835..8867
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10015..10510

;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 12278..12443
; US-08-256-799-4

Query Match 78.1%; Score 16.4; DB 4; Length 13104;
Best Local Similarity 94.4%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gggtggggaggctatttgg 21
||||| |||||||
Db 10242 GGGTGGGTGGCTATTGG 10225

RESULT 10
US-08-462-437-4/c
; Sequence 4, Application US/08462437
; Patent No. 6232094
; GENERAL INFORMATION:
; APPLICANT: HANSSON, Lennart
; APPLICANT: STROEMQVIST, Mats
; APPLICANT: BERGSTROEM, Sven
; APPLICANT: HERNELL, Olle
; APPLICANT: TOERNELL, Jan
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS
; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,437
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 88/92
; FILING DATE: 23-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: HANSSON-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELETYPE: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: ps459
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..8834
; FEATURE:
; NAME/KEY: intron
; LOCATION: 8868..10014
; FEATURE:
; NAME/KEY: intron
```

; LOCATION: 10511..12277  
; FEATURE: exon  
; NAME/KEY: 8835..8867  
; LOCATION: 8835..8867  
; FEATURE: exon  
; NAME/KEY: 10015..10510  
; LOCATION: 10015..10510  
; NAME/KEY: 12278..12443  
; LOCATION: 12278..12443  
US-08-462-437-4

Query Match 78.1%; Score 16.4; DB 4; Length 13104;  
Best Local Similarity 94.4%; Pred. No. 14;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggggtggaggctatttgg 21  
||||| |||||||  
Db 10242 GGGTGGGCTATTGG 10225

RESULT 11  
US-09-111-470-7  
; Sequence 7, Application US/09111470  
; Patent No. 6277959  
; GENERAL INFORMATION:  
; APPLICANT: Valladeau, Jenny  
; APPLICANT: Ravel, Odile  
; APPLICANT: Bates, Elizabeth E.M.  
; APPLICANT: Ford, John  
; APPLICANT: Saeland, Sem  
; APPLICANT: Lebecque, Serge J.E.  
; TITLE OF INVENTION: Mammalian Membrane Protein Genes;  
; TITLE OF INVENTION: Related Reagents  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/111.470  
; FILING DATE: 08-JUL-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/053,080  
; FILING DATE: 09-JUL-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: SF0695  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650)852-9196  
; TELEFAX: (650)496-1200  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1418 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 279..992  
; FEATURE:

; NAME/KEY: misc\_feature  
; LOCATION: 1348  
; OTHER INFORMATION: /note= "poly-A addition motif"  
US-09-111-470-7

Query Match 77.1%; Score 16.2; DB 4; Length 1418;  
Best Local Similarity 85.7%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21  
||||| |||||||  
Db 602 AAAGGATTGGAGGCTATTGG 622

RESULT 12  
US-08-147-777-3  
; Sequence 3, Application US/08147777  
; Patent No. 5914265  
; GENERAL INFORMATION:  
; APPLICANT: Roop, Dennis R.  
; APPLICANT: Rothnagel, Joseph A.  
; APPLICANT: Greenhalgh, David A.  
; APPLICANT: Yuspa, Stuart H.  
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LYON & LYON  
; STREET: 611 West Sixth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/147,777  
; FILING DATE:

; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: described below:  
; APPLICATION NUMBER: 07/876,289  
; FILING DATE: April 30, 1992  
; APPLICATION NUMBER: Unassigned (204/144)  
; FILING DATE: October 29, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 204/153  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 24979 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-147-777-3

Query Match 77.1%; Score 16.2; DB 2; Length 24979;  
Best Local Similarity 85.7%; Pred. No. 19;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21





;  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/136,277  
; FILING DATE: 15-OCT-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PATCH, Andrew J  
; REGISTRATION NUMBER: 32,925  
; REFERENCE/DOCKET NUMBER: B2272  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-521-2297  
; TELEFAX: 703-685-0573  
; TELEX: 248425 EMBON  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1025 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-136-277-23

Query Match 75.2%; Score 15.8; DB 1; Length 1025;  
Best Local Similarity 89.5%; Pred. No. 20;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 taggggtggaggtatttg 20  
|||||  
Db 148 TAGGTTGGGAGGCTATGTG 166

Search completed: February 15, 2002, 19:07:24  
Job time: 14618 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:03:38 ; Search time 9904.61 seconds  
(without alignments)  
22.783 Million cell updates/sec

Title: US-09-698-903B-3

Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
c 1	17.4	82.9	212	11	BG983367
c 2	17.4	82.9	429	11	BF924973
c 3	17	81.0	501	10	AW638905
c 4	17	81.0	583	10	AW635936
c 5	16.8	80.0	285	10	AV145846
c 6	16.8	80.0	321	11	F05693
c 7	16.8	80.0	424	11	W57725
c 8	16.8	80.0	429	11	BI188834
c 9	16.8	80.0	587	11	BG494856
c 10	16.8	80.0	626	10	AW644184
c 11	16.8	80.0	667	10	AW639842
c 12	16.8	80.0	892	11	BF167952

c 13	16.8	80.0	914	13	CNS03N93
c 14	16.4	78.1	184	13	AZ016369
c 15	16.4	78.1	284	11	BG210286
c 16	16.4	78.1	385	10	AW736851
c 17	16.4	78.1	462	11	R41615
c 18	16.4	78.1	575	10	AW578936
c 19	16.4	78.1	658	13	BH036913
c 20	16.4	78.1	714	11	BG217189
c 21	16.4	78.1	728	11	BG204492
c 22	16.4	78.1	781	13	AZ188345
c 23	16.4	78.1	798	11	BG199886
c 24	16.4	78.1	822	11	BG194262
c 25	16.4	78.1	829	11	BG199887
c 26	16.4	78.1	837	11	BG211849
c 27	16.4	78.1	838	11	BG209779
c 28	16.4	78.1	850	11	BG215054
c 29	16.4	78.1	859	11	BG219740
c 30	16.4	78.1	866	11	BG205014
c 31	16.4	78.1	962	11	BG195797
c 32	16.2	77.1	162	10	BE011473
c 33	16.2	77.1	166	11	BF328785
c 34	16.2	77.1	228	10	AV338030
c 35	16.2	77.1	263	10	AA178060
c 36	16.2	77.1	265	10	BB252853
c 37	16.2	77.1	270	10	BB307020
c 38	16.2	77.1	274	10	BB008234
c 39	16.2	77.1	298	10	BB464468
c 40	16.2	77.1	321	10	AA850434
c 41	16.2	77.1	323	10	BB319109
c 42	16.2	77.1	326	10	BB319885
c 43	16.2	77.1	326	10	BB461095
c 44	16.2	77.1	328	13	AQ610088
c 45	16.2	77.1	385	10	BE113833

ALIGNMENTS

RESULT 1  
BG983367/c  
LOCUS BG983367 212 bp mRNA EST 12-JUN-2001  
DEFINITION IL5-CN0067-060301-384-fl2 CN0067 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG983367  
VERSION BG983367.1 GI:14386102  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 212)  
AUTHORS Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, C.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&tl2=IL5-CN0067-060301-384-fl2&tl3=2001-03-06&tl4=1)  
Seq primer: puc 18 forward

High quality sequence stop: 211.  
 Location/Qualifiers  
 1. .212  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="CN0067"  
 /dev\_stage="Adult"  
 /note="Organ: colon\_normal; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 38 a 84 c 32 g 58 t  
 ORIGIN

Query Match 82.9%; Score 17.4; DB 11; Length 212;  
 Best Local Similarity 94.7%; Pred. No. 6e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 taggggtggggagctatttg 20  
 |||||  
 Db 108 TAGGGTGGGAGGCTATG 90

RESULT 2  
 BF924973  
 LOCUS IL2-NT0200-281100-263-C11 NT0200 Homo sapiens cDNA, mRNA sequence. 19-JAN-2001  
 DEFINITION  
 ACCESSION BF924973  
 VERSION  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 429)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&t2=IL2-NT0200-  
 281100-263-C11&t3=2000-11-28&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 351.  
 Location/Qualifiers  
 1. .429  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NT0200"  
 /dev\_stage="Adult"  
 /note="Organ: nervous tumor; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 99 a 106 c 132 g 90 t 2 others  
 ORIGIN

Query Match 82.9%; Score 17.4; DB 11; Length 429;  
 Best Local Similarity 94.7%; Pred. No. 6.2e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 3 aggggtggggagctatttg 21  
 |||||  
 Db 392 AGCGTGGGAGGCTATTGG 410

RESULT 3  
 AW638905/c  
 LOCUS AW638905 501 bp mRNA EST 26-APR-2001  
 DEFINITION bl76b06.w1 Blackshear/Soares normalized Xenopus egg library Xenopus  
 laevis cDNA clone PBX0076B06 5', mRNA sequence.  
 ACCESSION AW638905  
 VERSION  
 KEYWORDS EST.  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 Xenopodinae; Xenopus.  
 1 (bases 1 to 501)  
 Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.,  
 Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman  
 ,J.W., Bonaldo,M.F. and Soares,M.B.  
 The NIEHS Xenopus maternal EST project: interim analysis of the  
 first 13,879 ESTs from unfertilized eggs  
 Gene 267 (1), 71-87 (2001)  
 21211403  
 Contact: Perry J. Blackshear  
 Office of Clinical Research and Laboratory of Signal Transduction  
 National Institute of Environmental Health Sciences  
 AL-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,  
 USA  
 Tel: 919 541-4899  
 Fax: 919 541-4571  
 Email: black009niehs.nih.gov  
 Clone is available through Research Genetics, Inc., 2130 Memorial  
 Parkway, Huntsville, AL 35901  
 Phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email  
 cdna@resgen.com  
 DNA Sequencing and analyses performed by National Institutes of  
 Health Intramural Sequencing Center (NISC).  
 PCR Primers  
 FORWARD: TGTAACACGACGCCAGT  
 BACKWARD: CAGGAACAGCTATGACC  
 Plate: 0076 row: B column: 06  
 Seq primer: T7 primer.  
 Location/Qualifiers  
 1. .501  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="PBX0076B06"  
 /clone\_lib="Blackshear/Soares normalized Xenopus egg  
 library"  
 /sex="female"  
 /tissue\_type="unfertilized egg"  
 /cell\_type="unfertilized egg"  
 /dev\_stage="unfertilized egg"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3-Pac; Site\_1: EcoRI; Site\_2: NotI;  
 PolyA-selected mRNA was prepared from unfertilized Xenopus  
 laevis eggs. The library was constructed in the vector  
 pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and  
 Soares, M.B. 'Normalization and subtraction: two

approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dt18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT73-Pac vector. The library contained approximately 7.2 x 10<sup>5</sup> recombinants, with average insert sizes of 1-1.5 kb."

# BASE COUNT ORIGIN

161 a 116 c 106 g 118 t

Query Match 81.0%; Score 17; DB 10; Length 501;

Best Local Similarity 100.0%; Pred. No. 9.3e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tagggtggaggctatt 18

|||||

Db 317 TAGGGTGGAGGCTATT 301

# RESULT 4 AW635936/c

LOCUS AW635936 583 bp mRNA EST 26-APR-2001  
DEFINITION bl40003.w1 blackshear/Soares normalized Xenopus egg library Xenopus laevis cDNA clone PBX0040D03 5', mRNA sequence.

ACCESSION AW635936

VERSION AW635936.1 GI:7393017

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 583)

REFERENCE Blackshear P.J., Lai W.S., Thorn J.M., Kennington, E.A., Staffa, N.G.

Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman

, J.W., Bonaldo, M.F. and Soares, M.B.

The NIEHS Xenopus maternal EST project: Interim analysis of the

first 13,879 ESTs from unfertilized eggs

Gene 267 (1), 71-87 (2001)

21211403

CONTACT: Perry J. Blackshear

Office of Clinical Research and Laboratory of Signal Transduction

National Institute of Environmental Health Sciences

A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,

USA

Tel: 919 541-4899

Fax: 919 541-4571

Email: black009@niehs.nih.gov

Clone is available through Research Genetics, Inc., 2130 Memorial

Parkway, Huntsville, AL 35901

phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email

cdna@resgen.com

DNA Sequencing and analyses performed by National Institutes of

Health Intramural Sequencing Center (NISC).

PCR Primers

FORWARD: TGTAACGACGGCCAGT

BACKWARD: CAGGAACAGCTATGACC

Plate: 0040 row: D column: 03

Seq primer: T7 primer.

Location/Qualifiers

1. 583

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone="PBX0040D03"

/clone\_lib="Blackshear/Soares normalized Xenopus egg

library"

/sex="female"

/tissue\_type="unfertilized egg"

/cell\_type="unfertilized egg"

/dev\_stage="unfertilized egg"

/lab\_host="DH10B"

/note="Vector: pT73-Pac; Site\_1: EcoRI; Site\_2: NotI;

PolyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT73-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dt18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT73-Pac vector. The library contained approximately 7.2 x 10<sup>5</sup> recombinants, with average insert sizes of 1-1.5 kb."

# BASE COUNT ORIGIN

185 a 126 c 128 g 144 t

Query Match 81.0%; Score 17; DB 10; Length 583;

Best Local Similarity 100.0%; Pred. No. 9.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tagggtggaggctatt 18

|||||

Db 433 TAGGGTGGAGGCTATT 417

# RESULT 5

LOCUS AV145846/c

DEFINITION AV145846 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA clone 2810449H10, mRNA sequence.

ACCESSION AV145846

VERSION AV145846.1 GI:5349979

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 285)

REFERENCE Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,

Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara

, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,

Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,

Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara

, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,

Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,

Muranatsu, M., Okazaki, Y. and Hayashizaki, Y.

RIKEN Mouse ESTs

Unpublished (1999)

CONTACT: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-resetc.riken.go.jp

Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

further details.

Location/Qualifiers

1. 285

/organism="Mus musculus"

/db\_xref="C57BL/6J"

/db\_xref="taxon:10090"

/clone="2810449H10"

/clone\_lib="Mus musculus C57BL/6J 10-11 day embryo"

/sex="mixed"

/dev\_stage="10-11 day embryo"

63 a 87 c 53 g 82 t

BASE COUNT

ORIGIN

```

Query Match      80.0%; Score 16.8; DB 10; Length 285;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tagggtagggagcatttgg 21
|||||
DB 47 TAGGGTGGGAGGCAATTAGG 28

RESULT 6
F05693/c
LOCUS
DEFINITION HSC0JC091 normalized infant brain cDNA Homo sapiens cDNA clone
C-0JC09, mRNA sequence.
ACCESSION F05693
VERSION F05693.1 GI:669509
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 321)
AUTHORS Auffray C., Behar G., Bois F., Bouchier C., da Silva C., Devignes
M.D., Duprat S., Houlgatte R., Jumeau M.N., Lamy B., Lorenzo F.,
Mitchell H., Mariage-Samson R., Pietu G., Pouliot Y.,
Sebastiani-Kabekthchis C. and Tessier A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MIDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-0jc09
Seq primer: (-21)M13 universal.
Location/Qualifiers
1. .321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-0jc09"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/organ="brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
Isolate=muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soaress, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"
BASE COUNT 108 a 53 c 46 g 114 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 11; Length 321;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tagggtagggagcatttgg 21
|||||
DB 106 TGGGGTGGGAGGCAATTGG 87

RESULT 7
W57725/c
LOCUS
DEFINITION zd20c10.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:341202 3', mRNA sequence.
ACCESSION W57725
VERSION W57725.1 GI:1364441
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman
M., Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,
Rifkin J., Rohlfing T., Soares M., Tan F., Trevaskis E., Waterston
R., Williamson A., Wohldmann P. and Wilson R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
MIDLINE Contact: Wilson RK
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 866 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 376.
Location/Qualifiers
1. .424
/organism="Homo sapiens"
/db_xref="GDB:1266577"
/db_xref="taxon:9606"
/clone="IMAGE:341202"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/organ="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTGTGTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."
BASE COUNT 145 a 87 c 77 g 115 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 11; Length 424;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atagggtggaggcatttgg 20
|||||
DB 374 ATGGGTGGGAGGCTTTG 355

RESULT 8
BI188834/c
LOCUS
DEFINITION BI188834 429 bp mRNA EST 10-JUL-2001
IMAGE:3412fs.f1 Fusarium sporotrichioides Tri 10 overexpressed cDNA
library Fusarium sporotrichioides cDNA clone d3d12fs 3', mRNA
sequence.
ACCESSION BI188834
VERSION BI188834.1 GI:14662513
KEYWORDS EST.
SOURCE Fusarium sporotrichioides.

```

ORGANISM Fusarium sporotrichioides  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreales; mitosporic Hypocreales; Fusarium.  
REFERENCE 1 (bases 1 to 429)  
AUTHORS Ren.Q., Tag.A., Peplow.A., Lai.H., Kupfer.C., Peterson.A., Beremand  
.M. and Roe.B.  
TITLE Analysis of a Fusarium sporotrichioides EST database  
JOURNAL Unpublished (2001)  
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
Department of Chemistry and Biochemistry  
Advanced Center for Genome Technology, University of Oklahoma  
620 Parrington Oval, Norman, OK 73019, USA  
Tel: 405 325 4912  
Fax: 405 325 7762  
Email: broe@ou.edu  
Contact Dr. Marian Beremand regarding clone availability Included  
is the best homolog from a blastx search of Genbank nr 04-09-01  
71 3.9 g1350690|emb|CAA09 (AJ010199) variable surface  
glycoprotein[  
Seq primer: M13-20  
High quality sequence stop: 405.  
Location/Qualifiers

## FEATURES

source

1. .429  
/organism="Fusarium sporotrichioides"  
/strain="Tri 10"  
/db\_xref="taxon:5514"  
/clone="d3d12fs"  
/clone\_lib="Fusarium sporotrichioides Tri 10 overexpressed  
cDNA library"  
/note="Vector: pBlueScript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript  
; 3' end of cDNA cloned into XhoI site of pBluescript"  
BASE COUNT 137 a 102 c 68 g 122 t  
ORIGIN

Query Match 80.0%; Score 16.8; DB 11; Length 429;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 atagggtggaggctatttg 20  
||| ||||| ||||| |||||  
Db 164 ATTGGGAGGAGGCTATTG 145

RESULT 9  
BG494856/c  
LOCUS BG494856 587 bp mRNA EST 27-MAR-2001  
DEFINITION 602540862F1 NIH\_MGC\_59 Homo sapiens cDNA clone IMAGE:4671903 5',  
mRNA sequence.  
ACCESSION BG494856  
VERSION BG494856.1 GI:13456371  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 587)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Ph.D.  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: CLONETECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1487 row: 1 column: 16  
High quality sequence stop: 584.  
Location/Qualifiers

## FEATURES

Location/Qualifiers

source 1. .587  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4671903"  
/clone\_lib="NIH\_MGC\_59"  
/tissue\_type="mucoepidermoid carcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTATGACC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGGCCGACATG-dT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
Library."  
BASE COUNT 197 a 104 c 87 g 199 t  
ORIGIN

Query Match 80.0%; Score 16.8; DB 11; Length 587;  
Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 tagggtggaggctatttg 21  
||||| ||||| ||||| |||||  
Db 191 TGGGTGGAGGCAATTGG 172

RESULT 10  
AW644184  
LOCUS AW644184 626 bp mRNA EST 26-APR-2001  
DEFINITION CM38402.w1 Blackshear/Soares normalized Xenopus egg library Xenopus  
laevis cDNA clone PBX0137D02 5', mRNA sequence.  
ACCESSION AW644184  
VERSION AW644184.1 GI:7401561  
KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 626)  
AUTHORS Blackshear P.J., Lai W.S., Thorn J.M., Kennington E.A., Staffa N.G.  
Jr., Moore D.T., Bouffard G.G., Beckstrom-Sternberg S.M., Touchman  
J.W., Bonaldo M.F. and Soares M.B.  
TITLE The NIHs Xenopus maternal EST project: Interim analysis of the  
first 13,879 ESTs from unfertilized eggs  
JOURNAL Gene 267 (1), 71-87 (2001)  
MEDLINE 21211403  
COMMENT Contact: Perry J. Blackshear  
Office of Clinical Research and Laboratory of Signal Transduction  
National Institute of Environmental Health Sciences  
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,  
USA  
Tel: 919 541-4899  
Fax: 919 541-4571  
Email: black009@niehs.nih.gov  
Clone is available through Research Genetics, Inc., 2130 Memorial  
Parkway, Huntsville, AL 35901  
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email  
cdna@resgen.com  
DNA Sequencing and analyses performed by National Institutes of  
Health Intramural Sequencing Center (NISC).  
PCR Primers  
FORWARD: TGTAAACGACGCCAGT  
BACKWARD: CAGGAACAGCTATGACC  
Plate: 0137 row: D column: 02  
Seq primer: T7 primer.  
Location/Qualifiers

```

BACKWARD: CAGGAACAGCTATGACC
Plate: 0087 row: H column: 11
Seq primer: r7 primer.
Location/Qualifiers
1..567
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0087H11"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/note="Vector: p7T73-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
p7T73-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dm18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested p7T73-Pac vector.
The library contained approximately 7.2 X 10^5
recombinants, with average insert sizes of 1-1.5 kb."
214 a 124 c 144 g 185 t

```

BASE COUNT  
ORIGIN

Query Match	80.0%;	Score 16.8;	DB 10;	Length 567;
Best Local Similarity	90.0%;	Pred. No. 1.2e+03;		

Qy	2	taggtgggagggctatttg	21
Dd	143	TAGGCTGGGAGGATATTAGG	162

Accession: BF167952  
Version: BF167952.1  
Keywords: EST, house mouse, mRNA sequence.  
Source: IMAGE:401644 5'

JOURNAL COMMENT  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.llnl.gov>  
 Plate: L1AM9266 row: 0 column: 21  
 High quality sequence stop: 662.  
 Location/Qualifiers  
 1. .892  
   /organism="Mus musculus"  
   /strain="CZECH II (feral)"  
   /db\_xref="taxon:10090"  
   /clone="IMAGE:4017644"



```

/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT      241 a   225 c   237 g   189 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 11; Length 892;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2 tagggtagggagcgtattgg 21
||||| ||||| ||||| |||||
Db  871 TAGGCTGTGAGGCTATATGG 852

RESULT 13
CNS03N93/c      914 bp      DNA      GSS      17-MAY-2000
LOCUS      Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION  039K16 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION  AL251760
VERSION    AL251760.1 GI:7972772
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 914)
AUTHORS   Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE     Freshwater pufferfish Tetraodon nigroviridis
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 914)
AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernet,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE     Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL   Unpublished
REFERENCE  3 (bases 1 to 914)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT   This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
    source
        Location/Qualifiers
            1..914
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="039K16"
                /clone_lib="G"
                /note="Genoscope sequence ID : CORG039BF08Lp1-end : T7"

BASE COUNT      202 a   257 c   239 g   212 t      4 others
ORIGIN

Query Match      80.0%; Score 16.8; DB 13; Length 914;
Best Local Similarity 85.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

/clone_lib="Mus musculus"
/tissue_type="muscle"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI MethyIase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT      40 a   36 c   49 g   59 t
ORIGIN

Query Match      78.1%; Score 16.4; DB 13; Length 184;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  3 agggtagggagcgtattgg 20
||||| ||||| ||||| |||||
Db  60 AGGGTGGGAGGCAATTG 77

RESULT 15
BG210286/c      284 bp      mRNA      EST      21-APR-2001
LOCUS      BG210286
DEFINITION  RST29821 Athensys RAGE Library Homo sapiens cDNA, mRNA sequence.

```

```

Qy  1 ataggtagggagcgtattgg 21
||||| ||||| ||||| |||||
Db  281 ATGGGTGGGAGCNACTTGG 261

RESULT 14
AZ016369      184 bp      DNA      GSS      25-FEB-2000
LOCUS      RPCI-23-264M11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-264M11
DEFINITION  , DNA sequence.
ACCESSION  AZ016369
VERSION    AZ016369.1 GI:7091753
KEYWORDS   GSS.
SOURCE     Mus musculus.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 184)
AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE     Mouse BAC End Sequences from Library RPCI-23
JOURNAL   Unpublished (1999)
COMMENT   Other_GSSs: RPCI-23-264M11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac-ends/mouse/bac_end_intro.html
Plate: 264 row: M column: 11
Seq primer: SP6
Class: BAC ends.
FEATURES
    Location/Qualifiers
        1..184
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-264M11"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI MethyIase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT      40 a   36 c   49 g   59 t
ORIGIN

Query Match      78.1%; Score 16.4; DB 13; Length 184;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  3 agggtagggagcgtattgg 20
||||| ||||| ||||| |||||
Db  60 AGGGTGGGAGGCAATTG 77

RESULT 15
BG210286/c      284 bp      mRNA      EST      21-APR-2001
LOCUS      BG210286
DEFINITION  RST29821 Athensys RAGE Library Homo sapiens cDNA, mRNA sequence.

```

ACCESSION BG210286  
 VERSION BG210286.1 GI:13731973  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 284)  
 AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,  
 Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,  
 Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Mays, R., Smith, E.,  
 Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,  
 and Ducar, M.  
 TITLE Creation of Genome-wide Protein Expression Libraries using Random  
 Activation of Gene Expression  
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press  
 COMMENT Contact: Scott J. Cain  
 Athersys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596  
 Email: scai@atersys.com  
 High quality sequence stop: 234.  
 FEATURES  
 Location/Qualifiers  
 1..284  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Athersys RAGE Library"  
 /cell\_line="Hr1080"  
 /note="See 'Creation of Genome-wide Protein Expression  
 Libraries using Random Activation of Gene Expression',  
 Nature Biotechnology, in press. Note that even though the  
 cell type indicated is Hr1080, since a random activation  
 method was used, these sequence tags are not necessarily  
 expressed in Hr1080 under normal circumstances."  
 BASE COUNT 88 a 82 c 43 g 71 t  
 ORIGIN  
 Query Match 78.1%; Score 16.4; DB 11; Length 284;  
 Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 4 ggggtgggagcgtattgg 21  
 ||||| |||||  
 Db 238 GGGTGGGTGGCTATTGG 221

Search completed: February 15, 2002, 18:03:43  
 Job time: 20932 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:46:57 ; Search time 2553.1 Seconds  
(without alignments)  
96.924 Million cell updates/sec

Title: US-09-698-903b-4

Perfect score: 15  
Sequence: 1 ntcgastwtsgwgtt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_om.\*  
20: em\_or.\*  
21: em\_ov.\*  
22: em\_pat.\*  
23: em\_ph.\*  
24: em\_pl.\*  
25: em\_ro.\*  
26: em\_sts.\*  
27: em\_sy.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htgo\_hum.\*  
31: em\_htgo\_inv.\*  
32: em\_htgo\_rod.\*  
33: em\_htg\_hum.\*  
34: em\_htg\_inv.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	12.4	82.7	15	6	AX046662 Sequence
2	12.4	82.7	15	6	AX063588 Sequence
3	12.4	82.7	15	6	AX127751 Sequence
4	12.4	82.7	15	6	AX167666 Sequence
5	12.4	82.7	15	6	E08065 Non-specifi
6	12.4	82.7	206	10	MMU080260
7	12.4	82.7	240	8	GMU26701
8	12.4	82.7	411	8	PUMPAL2A
9	12.4	82.7	415	6	AX127755 Sequence
10	12.4	82.7	636	9	HSMCRP1
11	12.4	82.7	676	3	AF283269 Anopheles
12	12.4	82.7	733	3	GRO251757
13	12.4	82.7	905	8	CGU94186
14	12.4	82.7	927	8	AF049922
15	12.4	82.7	942	8	AF328994 Arabidops
16	12.4	82.7	1088	8	AF162910 Hirschfel
17	12.4	82.7	1092	8	AF161333 Raphanus
18	12.4	82.7	1092	8	AF162913 Raphanus
19	12.4	82.7	1321	6	AX046673 Sequence
20	12.4	82.7	1499	3	U24436
21	12.4	82.7	2115	8	CTR242496
22	12.4	82.7	2154	8	AF052582 Fusarium
23	12.4	82.7	2246	8	AF387019 Arabidops
24	12.4	82.7	2275	3	AF218866 Caenorhab
25	12.4	82.7	2384	8	PCPAL2
26	12.4	82.7	2427	9	AF191338 Homo sapi
27	12.4	82.7	2851	10	RATAP0A4G
28	12.4	82.7	2914	9	AK022864 Homo sapi
29	12.4	82.7	2931	8	SCYGR065C
30	12.4	82.7	3084	3	AF395331 Drosophil
31	12.4	82.7	3246	3	AF245516 Drosophil
32	12.4	82.7	3282	2	AC019765 Drosophil
33	12.4	82.7	3614	3	DMEI31630
34	12.4	82.7	3855	10	BC005506 Mus muscu
35	12.4	82.7	3887	3	AF185069 Leishmani
36	12.4	82.7	4233	6	AX017147 Sequence
37	12.4	82.7	4271	9	AB029020 Homo sapi
38	12.4	82.7	4501	14	MVU88057
39	12.4	82.7	4953	4	RABGLCOGEN
40	12.4	82.7	5288	2	AC019630
41	12.4	82.7	5315	2	WHTGIR
42	12.4	82.7	6201	2	AC018291 Drosophil
43	12.4	82.7	6211	3	AF325859 Choriston
44	12.4	82.7	7101	3	AF361883 Schistoso
45	12.4	82.7	9276	2	AC014293 Drosophil

ALIGNMENTS

RESULT 1					
AX046662	AX046662	Sequence 8 from Patent WO0068403.	15 bp	DNA	PAT 15-DEC-2000
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					

AX046662  
Sequence 8 from Patent WO0068403.  
AX046662.1 GI:11876228  
synthetic construct.  
synthetic construct  
artificial sequence.  
1 (bases 1 to 15)  
Paul.W.B., Scott,R.J., Hird,D.U. and Hodge,R.U.  
Tapetum-specific promoters  
Patent: WO 0068403-A 8 16-NOV-2000;  
BIOGENMA UK LIMITED (GB)  
Location/Qualifiers  
1..15  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer"

```

misc_feature 1
/note="n is a, t, g, or c"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGTT 15

RESULT 2
AX063588
LOCUS AX063588 15 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 7 from Patent WO0100801.
ACCESSION AX063588
VERSION AX063588.1 GI:12541312
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 15)
AUTHORS Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 7 04-JAN-2001;
Novartis AG (CH); Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic Oligonucleotide"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGTT 15

RESULT 3
AX127751
LOCUS AX127751 15 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 4 from Patent WO0131042.
ACCESSION AX127751
VERSION AX127751.1 GI:14134398
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 15)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 4 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer MDB285"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

```

```

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGTT 15

RESULT 4
AX167666
LOCUS AX167666 15 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 11 from Patent WO0144277.
ACCESSION AX167666
VERSION AX167666.1 GI:14597053
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 15)
AUTHORS wegrich Glover,L., Budziszewski,G.J., Levin,J.Z. and Zhou,Q.
TITLE Herbicide target genes and methods
JOURNAL Patent: WO 0144277-A 11 21-JUN-2001;
Syngenta Participations AG (CH)
FEATURES
source Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="oligonucleotide"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGTT 15

RESULT 5
E08065
LOCUS E08065 15 bp DNA PAT 29-SEP-1997
DEFINITION Non-specific primer.
ACCESSION E08065
VERSION E08065.1 GI:2176190
KEYWORDS JP 1994253843-A/4.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 15)
AUTHORS Riyuu,Y., Robaato,E.U. and MitsuKawa,N.
TITLE METHOD FOR AMPLIFYING DNA SEQUENCE
JOURNAL Patent: JP 1994253843-A 4 13-SEP-1994;
CHIKUBU KANKYO SANGYO GIJUTSU KENKYU KIKO, MITSUI GIYOUSAI
SHOKUBUTSU BIO KENKYUSHO:KK
COMMENT
OS None
OC Artificial sequences.
PN JP 1994253843-A/4
PD 13-SEP-1994
PF 03-MAR-1993 JP 1993043006
PI RIYUU YOKUOU, ROBAATO EFU UITSUTEI, MITSUKAWA NORIHIRO PC
C12N15/10,C12Q1/68;
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FH source 1..15
FT /organism='Artificial sequences'

```

FEATURES  
source Location/Qualifiers  
1. .15  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 1 a 1 c 3 g 5 t 5 others  
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||:|:|:|:|:|:|  
Db 2 TCGASTWTSGWGTT 15

RESULT 6  
MMU80260 206 bp DNA ROD 19-DEC-1996  
LOCUS Mus musculus clone ma9 L1 retrotransposon LINE1 repeat region.  
DEFINITION U80260  
ACCESSION U80260  
VERSION U80260.1 GI:1737239  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 206)  
AUTHORS Woodcock,D.M. and Linsemeyer,M.E.  
TITLE Direct Submission  
JOURNAL Submitted (29-NOV-1996) Research, Peter MacCallum Cancer Institute, St. Andrew's Place, East Melbourne, Vic 3000, Australia  
FEATURES  
source Location/Qualifiers  
1. .206  
/organism="Mus musculus"  
/transposon="L1 retrotransposon"  
/db\_xref="taxon:10090"  
/clone="ma9"  
repeat\_region <1..>206  
BASE COUNT 32 a 14 c 74 g 86 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 206;  
Best Local Similarity 71.4%; Pred. No. 2e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||:|:|:|:|:|  
Db 45 TCGAGTTTCGAGTT 58

RESULT 7  
GMU26701/c 240 bp DNA PLN 26-JUL-2001  
LOCUS Glycine max satellite STR120-B.1.  
DEFINITION U26701  
ACCESSION U26701  
VERSION U26701.1 GI:1142703  
KEYWORDS soybean.  
SOURCE Glycine max  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
REFERENCE 1 (bases 1 to 240)  
AUTHORS Morgante,M., Jurman,I., Shi,L., Zhu,T., Keim,P. and Rafalski,J.A.  
TITLE The STR120 satellite DNA of soybean: organization, evolution and chromosomal specificity  
JOURNAL Chromosome Res. 5 (6), 363-373 (1997)  
MEDLINE 98030220

PUBMED 9364938  
REFERENCE 2 (bases 1 to 240)  
AUTHORS Morgante,M. and Rafalski,J.A.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-1995) Michele Morgante, Dip. di Produzione Vegetale, Universita di Udine I-33100, Via delle Scienze, 208, udine I-33100, Italy  
FEATURES  
source Location/Qualifiers  
1. .240  
/organism="Glycine max"  
/cultivar="Ripley"  
/db\_xref="taxon:3847"  
repeat\_unit 1. .120  
satellite 1. .240  
/note="STR120-B.1"  
repeat\_unit 121. 240  
BASE COUNT 81 a 49 c 60 g 50 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 8; Length 240;  
Best Local Similarity 71.4%; Pred. No. 2e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||:|:|:|:|:|  
Db 132 TCGAGTTTCGAGTT 119

RESULT 8  
PUMPAL2A 411 bp DNA PLN 21-JUL-1995  
LOCUS Petroselinum crispum phenylalanine ammonia-lyase (PAL-2) gene,  
DEFINITION promoter region and 5' end of cds.  
ACCESSION L37355  
VERSION L37355.1 GI:567856  
KEYWORDS phenylalanine ammonia-lyase.  
SOURCE Petroselinum crispum DNA.  
ORGANISM Petroselinum crispum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum.  
REFERENCE 1 (bases 1 to 411)  
AUTHORS Loemann,E., Parniske,M. and Hahlbrock,K.  
TITLE Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92 (13), 5905-5909 (1995)  
MEDLINE 95320184  
FEATURES  
source Location/Qualifiers  
1. .411  
/organism="Petroselinum crispum"  
/db\_xref="taxon:4043"  
/note="(vector lambda EMBL4)"  
promoter 1. .260  
TATA\_signal 230. .235  
gene 409. .411  
/gene="PAL-2"  
BASE COUNT 128 a 99 c 40 g 144 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 8; Length 411;  
Best Local Similarity 71.4%; Pred. No. 2e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||:|:|:|:|:|  
Db 363 TCGAGTTTCGAGTT 376

RESULT 9  
AX127755 415 bp DNA PAT 15-MAY-2001  
LOCUS AX127755

```

DEFINITION Sequence 8 from Patent WO0131042.
ACCESSION AX127755
VERSION AX127755.1 GI:14134402
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 415)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 8 03-MAY-2001,
Aventis CropScience N.V. (BE)
FEATURES
    source
        1..415
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="5' border flanking region of elite event MS-B2"
        misc_feature
            1..234
                /note="plant DNA"
        misc_feature
            235..415
                /note="T-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN
    1..415
        tcgastwtsgwgtt 15
        |||||:|:|:|
        2 TCGAGTTTGGTGT 15

Query Match 82.7%; Score 12.4; DB 6; Length 415;
Best Local Similarity 71.4%; Pred. No. 2e+03; 0; Indels 0; Gaps 0;
Matches 10; Conservative 4; Mismatches 0;

QY 2 tcgastwtsgwgtt 15
Db 2 TCGAGTTTGGTGT 15

RESULT 10
HSMCRP1/c 636 bp DNA PRI 11-NOV-1998
LOCUS Homo sapiens MHC class I-related protein MRI precursor (MRI) gene,
DEFINITION signal peptide.
ACCESSION AF073484
VERSION AF073484.1 GI:3859892
KEYWORDS I of 2
SEGMENT human.
SOURCE
    ORGANISM
        Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 636)
AUTHORS Yamaguchi,H., Kurosawa,Y. and Hashimoto,K.
TITLE Expanded genomic organization of conserved mammalian MHC class
I-related genes, human MRI and its murine ortholog
JOURNAL Biochem. Biophys. Res. Commun. 250 (3), 558-564 (1998)
MEDLINE 99003494
REFERENCE 2 (bases 1 to 636)
AUTHORS Hashimoto,K.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1998) Institute for Comprehensive Medical
Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan
FEATURES
    source
        1..636
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="placenta"
            /chromosome="1"
            /map="1q25.3"
            /map="1q25.3"
            /gene="MRI"
BASE COUNT 160 a 140 c 138 g 198 t
ORIGIN
    1..636
        sig_peptide
            <1..45
            /gene="MRI"
BASE COUNT 160 a 140 c 138 g 198 t
ORIGIN

Query Match 82.7%; Score 12.4; DB 9; Length 636;
Best Local Similarity 71.4%; Pred. No. 2e+03;

QY 2 tcgastwtsgwgtt 15
Db 2 TCGAGTTTGGTGT 15

RESULT 12
GRO251757/c 733 bp mRNA INV 12-JAN-2001
LOCUS Globodera rostochiensis mRNA for for hypothetical protein (clone
DEFINITION A18);
ACCESSION AJ251757
VERSION AJ251757.1 GI:12227170
KEYWORDS ORF.
SOURCE Globodera rostochiensis.

```

---

```

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15
Db 88 TCGACTTTGGAGTT 75

RESULT 11
AF283269 676 bp mRNA INV 16-OCT-2000
LOCUS Anopheles gambiae ribosomal protein S26 (Irps26) mRNA, complete
DEFINITION cds
ACCESSION AF283269
VERSION AF283269.1 GI:10242303
KEYWORDS African malaria mosquito.
SOURCE Anopheles gambiae
ORGANISM
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
    Culicoides; Anopheles.
REFERENCE 1 (bases 1 to 676)
AUTHORS Oduol,F., Xu,J., Niare,O., Natarajan,R. and Vernick,K.D.
TITLE Genes identified by an expression screen of the vector mosquito
Anopheles gambiae display differential molecular immune response to
malaria parasites and bacteria
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11397-11402 (2000)
PUBMED 11005829
REFERENCE 2 (bases 1 to 676)
AUTHORS Oduol,F.O., Xu,J., Niare,O., Natarajan,R. and Vernick,K.D.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2000) Department of Medical and Molecular
Parasitology, New York University School of Medicine, 341 East 25th
Street, New York, NY 10010, USA
FEATURES
    source
        1..676
            /organism="Anopheles gambiae"
            /strain="G3"
            /db_xref="taxon:7165"
            /dev_stage="adult"
            1..676
            /gene="Irps26"
            /gene="Irps26"
            /note="immune-responsive"
            /codon_start=1
            /product="ribosomal protein S26"
            /protein_id="AAG15374.1"
            /db_xref="GI:10242304"
            /translation="MLERRNGGRCKHNRGHVAVRCTNCARCVPKDKAIKKFVIRNIV
            EAAAVRDISDASVYSVYLPKLYAKLHYCVSKAIHKSIVVNRSKETRRTTPQRSFP
            KDMNRQNAQRK"
BASE COUNT 160 a 179 c 195 g 142 t
ORIGIN

Query Match 82.7%; Score 12.4; DB 3; Length 676;
Best Local Similarity 71.4%; Pred. No. 2e+03; 0; Indels 0; Gaps 0;
Matches 10; Conservative 4; Mismatches 0;

QY 2 tcgastwtsgwgtt 15
Db 582 TCGAGTTTCGTGT 595

RESULT 12
GRO251757/c 733 bp mRNA INV 12-JAN-2001
LOCUS Globodera rostochiensis mRNA for for hypothetical protein (clone
DEFINITION A18);
ACCESSION AJ251757
VERSION AJ251757.1 GI:12227170
KEYWORDS ORF.
SOURCE Globodera rostochiensis.

```

```
ORGANISM Globodera rostochiensis
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidae; Heteroderidae; Heteroderinae; Globodera.
REFERENCE
AUTHORS Qln,L., Overmars,H., Helder,J., Popeijus,H., van der Voort,J.R.,
Groenink,W., van Keert,P., Schots,A., Bakker,J. and Smant,G.
TITLE An efficient cDNA-AFLP-based strategy for the identification of
putative pathogenicity factors from the potato cyst nematode
Globodera rostochiensis
JOURNAL Mol. Plant Microbe Interact. 13 (8), 830-836 (2000)
MEDLINE 20393248
REFERENCE 2 (bases 1 to 733)
AUTHORS Qln,L.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Qln L., Laboratory of Nematology, Wageningen
University, Binnenhaven 10, 6709 PD Wageningen, NETHERLANDS
FEATURES
source
1. 733
/organism="Globodera rostochiensis"
/isolate="Rol-Mierenbos"
/db_xref="taxon:31243"
/clone="A18"
/tissue_type="dorsal gland"
/dev_stage="infective second stage juvenile"
sig_peptide 23. .673
CDS 23. .673
/note="ORF"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC21848.1"
/db_xref="GI:12227171"
/translation="MPTFLFLAAVCLLVADASPPPTNTKLEKGPASAGNAESTPGLP
PQNWDYDAHKDLTLTDGLIVQNNKDKWRYVFAKOSIPKDGIFYEYVKILKFKGR
VOIGLATQKPLDKPVGSECTVAYASAGYFWGHKDSKCTIGTANGRPYIKGPWFNGS
DVICGVDFANROIFETKNGORLITTELHVDAAKLPCVSMFPGTKIEANFGPF"
BASE COUNT 210 a 168 c 171 g 184 t
ORIGIN
Query Match 82.7%; Score 12.4; DB 3; Length 733;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Qy 2 tcgastwtsgwgtt 15
||||:|:|:|
Db 105 TCGACTTGGGTGTT 92
RESULT 13
GCU94186 905 bp mRNA PLN 20-DEC-1997
LOCUS Colletotrichum gloeosporioides nitrogen starvation-induced
DEFINITION Colletotrichum gloeosporioides nitrogen starvation-induced
ACCESSION U94186
VERSION U94186.1 GI:2707190
KEYWORDS Glomerella cingulata.
SOURCE Glomerella cingulata.
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Glomerella.
REFERENCE 1 (bases 1 to 905)
AUTHORS Stephenson,S.-A., Maclean,D.J. and Manners,J.M.
TITLE Disruption of a novel pathogenicity gene of Colletotrichum
gloeosporioides results in a hypersensitive response in the host
Stylosanthes guianensis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 905)
AUTHORS Stephenson,S.-A., Maclean,D.J. and Manners,J.M.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1997) CRCTPP, University of Queensland, St.Lucia,
Brisbane, Queensland 4072, Australia
FEATURES
Location/Qualifiers
1. 905
/organism="Glomerella cingulata"
/isolate="UQ62"
/db_xref="taxon:5457"
/note="biotype B"
biotype: B"
74. .721
/note="glutamine rich protein similar to glutenins;
nitrogen starvation-induced"
/codon_start=1
/protein_id="AAB92223.1"
/db_xref="GI:2707191"
/translation="MSSNDYGNSGHQGYNQGYQQQGYGQGGYQPQPSYASQQQ
PQGYPQSHSPYQQGGYNAPQQHGYGCGPQPYAQGGNQGYSSGHDNRGHSYFQ
QQQYGGHQQQGYGQQGYGQQGYGQQGYGPPGPDGGERGLATLVGGGTAGWAAH
KAGGLGLTGLGAIAGAVGANMLPFAKHKHKKKKHKKGLSGSSSSSD"
BASE COUNT 233 a 273 c 243 g 156 t
ORIGIN
Query Match 82.7%; Score 12.4; DB 8; Length 905;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Qy 2 tcgastwtsgwgtt 15
||||:|:|:|
Db 29 TCGAGTATCGAGTT 16
RESULT 14
AF049922/c
LOCUS AF049922 927 bp mRNA PLN 05-JUL-2000
DEFINITION Petunia x hybrida PGPS/D6 (PGPS/D6) mRNA, complete cds.
ACCESSION AF049922
VERSION AF049922.1 GI:4105781
KEYWORDS Petunia x hybrida.
SOURCE Petunia x hybrida.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
REFERENCE 1 (bases 1 to 927)
AUTHORS Guyon,V.N., Astwood,J.D., Garner,E.C., Dunker,A.K. and Taylor,L.P.
TITLE Isolation and characterization of cDNAs expressed in the early
stages of flavonol-induced pollen germination in petunia
JOURNAL Plant Physiol. 123 (2), 699-710 (2000)
MEDLINE 20317212
PUBMED 10859200
REFERENCE 2 (bases 1 to 927)
AUTHORS Guyon,V., Astwood,J.D. and Taylor,L.P.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1998) Genetics and Cell Biology, Washington State
University, Science Hall, Pullman, WA 99164-4234, USA
FEATURES
source
1. 927
/organism="Petunia x hybrida"
/db_xref="taxon:4102"
/tissue_type="germinating petunia pollen treated with
kaempferol"
1. 927
/gene="PGPS/D6"
60. .707
/gene="PGPS/D6"
/note="ER lumen protein retaining receptor homolog;
putative HDEL receptor"
/codon_start=1
/product="PGPS/D6"
/protein_id="AAD02548.1"
/db_xref="GI:4105782"
/translation="MNIFRLAGDMTHLASVLVLLKTHITKSCAGVSLKQELVALVF
VTRYLDITFDISLYNTMKLVFLGSSLSIVYMRHHKIVRRSYDKDDQDFRHLFLVL
PCLLALVINEKFTFKVMTFTSYLEAVAILPOLVLLQRTNIDNTGQYIFLLGAY
RSFYILNWWYRYFTEPHVHWITWIAGLIQTLLYADFFYFYFQSWKNTKLELPA"
FEATURES
Location/Qualifiers
```

BASE COUNT 273 a 173 c 176 g 305 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 8; Length 927;  
Best Local Similarity 71.4%; Pred. No. 2e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
||||:|:|:|  
Db 694 TCGAGTTTCGTGTT 681

RESULT 15  
AF328994/c  
LOCUS  
DEFINITION Arabidopsis lyrata putative S-glycoprotein (S) gene, S-13-4 allele, partial cds. PLN 17-MAY-2001  
ACCESSION AF328994  
VERSION AF328994.1 GI:13345390  
KEYWORDS  
SOURCE Arabidopsis lyrata.  
ORGANISM Arabidopsis lyrata  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 942)  
Schierup,M.H., Mable,B.K., Awadalla,P. and Charlesworth,D.  
Identification and Characterization of a Polymorphic Receptor Kinase Gene Linked to the Self-Incompatibility Locus of Arabidopsis lyrata  
JOURNAL Genetics 158 (1), 387-399 (2001)  
MEDLINE 21231627  
PubMed 11333247  
REFERENCE 2 (bases 1 to 942)  
AUTHORS Schierup,M.H., Mable,B.K., Awadalla,P. and Charlesworth,D.  
TITLE Direct Submission  
JOURNAL Submitted (15-DEC-2000) Genetics and Ecology, University of Aarhus, Ny Munkegade, Building 540, Aarhus 8000, Denmark  
FEATURES  
source  
1..942  
/organism="Arabidopsis lyrata"  
/db\_xref="taxon:59689"  
<1..>942  
/gene="S"  
/product="putative S-glycoprotein"  
<1..>942  
/gene="S"  
/allele="S-13-4"  
<1..>942  
/gene="S"  
/codon\_start=3  
/product="putative S-glycoprotein"  
/protein\_id="AAK19315.1"  
/db\_xref="GI:13345391"  
/translation="GNLVILGHSDKLWSTNLTSGNVESLVVAELLANGNVFMYSSK  
DQGGFLWQSFDPDTLPQMKLGWDKTKLNRLRWKSSYDFSSGNFTYELETRGF  
PEEFLRTKDIPHRSGPDGIRISGPEERQVDIMVNFTEDEEVAYTFLVTHSIY  
SRLTMSYSGYQRTWTSIPAWGRQLWSAPMDLQCDLYPSCGPYAYCDTNPPLNCII  
RGFPWSEQNMMDGTSGCCVRRTPLSRRDGFLLPMKKMLPTTMTATVDRRISGKEC  
KOKCLMDCNCTAYANADIKNGLGCVIWTGELVDIRT"  
BASE COUNT 260 a 182 c 249 g 251 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 8; Length 942;  
Best Local Similarity 71.4%; Pred. No. 2e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
||||:|:|:|  
Db 302 TCGAGTTTCGAGTT 289





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:01:13 ; Search time 868.33 Seconds  
(without alignments)  
14.810 Million cell updates/sec

Title: US-09-698-903B-4

Perfect score: 15

Sequence: 1 ntcgastwtsgwgtt 15

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID22/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12.4	82.7	15	AAQ79175	Non-specific prime
2	12.4	82.7	15	AAA88455	PCR primer CA54 us
3	12.4	82.7	15	AAA48933	PCR primer MDB285
4	12.4	82.7	15	AAZ99438	PCR primer ARB1 fo
5	12.4	82.7	15	AAF90555	Degenerate primer
6	12.4	82.7	15	AAD06993	PCR primer MDB285
7	12.4	82.7	15	AAF66681	Degenerate TAIL-PC
8	12.4	82.7	15	AA89358	Arabidopsis DNA pr
9	12.4	82.7	15	AAC86474	Maize Mac2 gene de
c 10	12.4	82.7	30	AAC64167	Human pollinosis-a
11	12.4	82.7	304	AAH21085	Parallel detection

12	12.4	82.7	335	21	AAC02398	Human secreted pro
13	12.4	82.7	336	20	AAAX51486	Human secreted pro
c 14	12.4	82.7	388	20	AAAX20974	Polynucleotide seq
15	12.4	82.7	415	22	AA066997	Right (5') border
16	12.4	82.7	633	22	AAH21072	Parallel detection
17	12.4	82.7	679	22	AAH99881	Human protein enco
18	12.4	82.7	802	21	AAC64159	Human pollinosis-a
19	12.4	82.7	836	20	AAH81127	DNA sequence upstr
20	12.4	82.7	951	19	AAV03315	Pyrolabus fumarius
21	12.4	82.7	1321	22	AAC86489	Maize Mac2n-3 codi
22	12.4	82.7	2914	22	AAH17685	Human cDNA sequenc
23	12.4	82.7	3198	21	AAC48642	Arabidopsis thalia
24	12.4	82.7	4233	20	AAH87946	Candida tropicalis
25	12.4	82.7	4271	21	AAC64160	Human pollinosis-a
26	12.4	82.7	4345	22	AAH99730	Human protein enco
c 27	12.4	82.7	5098	20	AAH24984	E. coli MG1655 rrn
c 28	12.4	82.7	49272	19	AAV35000	Mycobacteriophage
29	11.4	76.0	45	20	AAZ24665	Oligonucleotide fo
c 30	11.4	76.0	165	22	AAI55830	Probe #24516 used
31	11.4	76.0	249	19	AAV03550	DNA sequence that
32	11.4	76.0	438	21	AAZ52235	Arabidopsis thalia
33	11.4	76.0	462	22	AAH52999	S. epidermidis ope
34	11.4	76.0	482	21	AAH38143	Zea mays DNA fragm
35	11.4	76.0	489	21	AAH79427	Eucalyptus grandis
c 36	11.4	76.0	503	21	AAH37052	Arabidopsis thalia
c 37	11.4	76.0	539	21	AAA45066	Human secreted exp
c 38	11.4	76.0	574	21	AAH09195	Fusarium venenatum
39	11.4	76.0	576	16	AAH083845	Hepatitis C virus
40	11.4	76.0	576	17	AAH16559	Hepatitis C virus
41	11.4	76.0	579	21	AAH10353	Fusarium venenatum
42	11.4	76.0	581	22	AAH29686	Drosophila melanog
c 43	11.4	76.0	590	22	AAI42770	Probe #11456 used
44	11.4	76.0	606	22	AAH07816	Human cDNA clone (
45	11.4	76.0	619	21	AAC48567	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAQ79175	AAQ79175 standard; DNA; 15 BP.
ID	AAQ79175;
AC	AAQ79175;
XX	
XX	
DT	22-JUN-1995 (first entry)
XX	
DE	Non-specific primer #1, for amplification of T-DNA.
XX	
KW	Polymerase chain reaction; PCR; amplify; primer; specific;
KW	non-specific; DNA polymerase; deoxyribonucleotide; cyclisation;
KW	ligation; ss.
XX	
OS	Synthetic.
XX	
PN	JP06253843-A.
XX	
PD	13-SEP-1994.
XX	
PF	03-MAR-1993; 93JP-0043006.
XX	
PR	03-MAR-1993; 93JP-0043006.
XX	
PA	(MITS-) MITSUI CYOSAI SHOKUBUTSU BIO KENKYUSHO.
PA	(CHIK-) ZH CHIKU KANKYO SANGYO GIJITSU KENKYU.
XX	
DR	WPI; 1994-328991/41.
XX	
PT	New DNA amplification method - without cyclisation of the DNA or
PT	ligation of oligonucleotide(s)
XX	
PS	Example 1; Page 10; 14pp; Japanese.
XX	

CC The sequences given in AA079172-74 are specific primers which were  
 CC used in conjunction with the non-specific primers given in AA079175-77  
 CC in the DNA amplification method of the invention. The method  
 CC comprises mixing a DNA sequence and a primer, DNA polymerase and  
 CC deoxyribonucleotides. The DNA is denatured allowing annealing of  
 CC the primers. The amplification cycle consists of one cycle at a  
 CC temperature to allow non-specific annealing of non-specific primers,  
 CC numerous cycles at a temp. to allow specific annealing of specific  
 CC primers and numerous cycles at a temp to allow non-specific primers  
 CC to hybridise specifically to complementary sequences. This method  
 CC allows amplification of neighbouring sequences with known sequences,  
 CC efficiently without cyclisation of the DNA or ligation of  
 CC oligonucleotides.  
 XX  
 SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 15; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 |||||  
 DB 2 tcgastwtsgwgtt 15

## RESULT 2

AAA88455  
 ID AAA88455 standard; DNA; 15 BP.

AC AAA88455;

XX 09-JAN-2001 (first entry)

DE PCR primer CA54 used to identify Arabidopsis 16713 gene.

KW 16713 gene; essential gene; herbicide; screening; transgenic plant;  
 Arabidopsis thaliana; PCR primer; ss.

OS Agrobacterium sp.

PN WO200053782-A2.

XX 14-SEP-2000.

XX 03-MAR-2000; 2000WO-EP01884.

XX 05-MAR-1999; 99US-0263385.

PR 10-MAY-1999; 99US-0309036.

PR 08-JUN-1999; 99US-0327662.

XX 29-SEP-1999; 99US-0408196.

PA (NOVS ) NOVARTIS AG.

XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Patton DA, Ashby CS, Thomas CR, Mcelver JA, Levin JZ;

PI Budziszewski GJ;

XX WPI; 2000-594330/56.

PT Novel genes isolated from Arabidopsis coding for proteins essential for  
 PT normal plant development, are useful to discover new herbicides and to  
 PT identify inhibitors that are potential herbicides -

PS Example 4c; Page 53; 92pp; English.

CC This degenerate oligonucleotide, termed CA54, is a primer used in  
 CC the border rescue of the 16713 gene (see AAA88448) from Arabidopsis  
 CC thaliana embryo-lethal line #16713. DNA flanking the borders of  
 CC line #16713 was isolated using TAIL PCR. CA54 is 1 of 6 primers  
 CC (see AAA88451-56) used in combination with 2 sets of nested, T-DNA  
 CC specific primers for the right border (see AAA88457-59) as well as  
 CC for the left border (see AAA88460-62) of the T-DNA region of pCSA104.

CC The 16713 gene encodes a protein (see AAB19528) that is essential for  
 CC normal plant development. Chemicals that inhibit the function of  
 CC the 16713-encoded protein are likely to have detrimental effects on  
 CC plants and are potentially good herbicide candidates.

XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 |||||  
 DB 2 tcgastwtsgwgtt 15

## RESULT 3

AAA48993

XX ID AAA48993 standard; DNA; 15 BP.

AC AAA48993;

XX 28-NOV-2000 (first entry)

DE PCR primer MDB285 used to locate bar gene insertion in rice plant.

KW Herbicide; phosphinothricin; glufosinate; bialaphos; bar;

KW phosphinothricin acetyl transferase; PAT; PCR primer; TAIL;

XX thermal asymmetric interlaced; rice; ss.

OS Synthetic.

XX WO200026345-A1.

XX 11-MAY-2000.

XX 03-NOV-1999; 99WO-US25667.

XX 03-NOV-1998; 98US-0185244.

XX (PLBZ ) PLANT GENETIC SYSTEMS NV.

XX Michiels F, Johnson K;

XX WPI; 2000-365598/31.

XX New glufosinate tolerant rice plants with genomic DNA capable of  
 XX yielding at least 3 restriction fragments and having bar gene under  
 XX control of CamV 35S promoter at specific location -

XX Example 3; Page 25; 44pp; English.

XX The present invention relates to a transgenic glufosinate tolerant rice  
 XX plant. The tolerant rice plants are generated by transformation with  
 XX the Streptomyces hygroscopicus bar gene. This gene encodes the enzyme  
 XX phosphinothricin acetyl transferase (PAT), which confers resistance to  
 XX the herbicidal compounds phosphinothricin (also called glufosinate) and  
 XX bialaphos. The 3' insertion site of the transgene was determined by  
 XX thermal asymmetric interlaced (TAIL) PCR. The present sequence is the  
 XX arbitrary degenerate PCR primer, MDB285, used for the TAIL PCR.  
 XX Glufosinate tolerant rice plants are useful as they give better yields  
 XX and growth, compared to untransformed plants, after application of a  
 XX glufosinate herbicide.

XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

Db 2 tcgastwtsgwgtt 15  
 |||

## RESULT 4

AAZ99438  
 ID AAZ99438 standard; DNA; 15 BP.

XX AAZ99438;

XX 03-JUL-2000 (first entry)

XX PCR primer ARB1 for the soybean AX5 promoter.

XX Gibberellic acid; copalyl diphosphate synthase; 3beta-hydroxylase;  
 KW 2-oxidase; phytoene synthase; C-20 oxidase; 2beta,3beta-hydroxylase;  
 KW seed germination; seedling growth; gibberellin biosynthetic pathway;  
 KW transgenic plant; hypocotyl; epicotyl; AX5 promoter; PCR primer; ss.  
 XX

OS Glycine max.

XX WO200009722-A2.

XX 24-FEB-2000.

XX 10-AUG-1999; 99WO-US18066.

XX 10-AUG-1998; 98US-0096111.

PR 07-JUN-1999; 99US-0137977.

XX (MONS ) MONSANTO CO.

XX Brown SM, Elich TD, Heck GR, Kishore GM, Logusch EW, Logusch SJ;  
 PI Pillier KJ, Rao S, Ream JE;

XX WPI; 2000-224351/19.

XX Obtaining transgenic plant useful for controlling seed germination and  
 PT seedling growth comprises transgene comprising a sequence expressing  
 PT altered levels of an essential hormone -

PS Example 7; Page 98; 267pp; English.

XX PCR primers AAZ99438-42 were used to amplify the soybean AX5 promoter.  
 CC The amplified sequence is used in the method of the invention. The  
 CC specification describes methods for the inhibition and control of  
 CC gibberellic acid levels. Gibberellic acid levels may be inhibited or  
 CC controlled by use of a chimeric expression construct expressing a RNA  
 CC or protein which suppresses the gibberellin biosynthetic pathway  
 CC sequence, diverts substrate from the pathway, or degrades pathway  
 CC substrates or products. The methods uses copalyl diphosphate synthase,  
 CC 3beta-hydroxylase, 2-oxidase, phytoene synthase, C-20 oxidase, and  
 CC a 2beta,3beta-hydroxylase polynucleotides to achieve this. The method  
 CC is used to control seed germination and seedling growth especially to  
 CC regulate gene products of gibberellin biosynthetic pathway and  
 CC restoration of normal seed germination, in transgenic plants. The  
 CC plants produced are gibberellin deficient, and have shortened hypocotyl  
 CC and/or epicotyl phenotypes compared to normal plants.

XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;

Best Local Similarity 100.0%; Pred. No. 3.1e+02; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

|||||

Db 2 tcgastwtsgwgtt 15

## RESULT 5

AAF90555

ID AAF90555 standard; DNA; 15 BP.

XX AAF90555;

XX 22-AUG-2001 (first entry)

XX Degenerate primer CA54 used to prime Arabidopsis genomic DNA.

XX ET1158 gene; GT6839 gene; ET5262 gene; herbicide; screening;  
 KW herbicide tolerance; transgenic plant; crop protection; PCR primer;  
 KW ss.

XX Synthetic.

XX WO200144277-A2.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000WO-EPI2748.

XX 16-DEC-1999; 99US-0465040.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Wegrich Glover L, Budziszewski GJ, Levin JZ, Zhou Q;

XX WPI; 2001-398122/42.

XX New herbicide target genes encoding proteins having ET1158, GT6839 or  
 PT ET5262 activity, for identifying an inhibitor of protein activity -

PS Example 1; Page 39; 67pp; English.

XX The present sequence is that of degenerate primer CA54. Arbitrary  
 CC degenerate primers LWAD1, CA51, CA52, CA53, CA54 and CA55 (see  
 CC AAF90551-56) were used to prime Arabidopsis thaliana genomic DNA  
 CC flanking a Ds transposon insertion. The degenerate primers were  
 CC used in combination with 2 sets of 3, nested, transposon-specific  
 CC primers (see AAF90557-62) that were homologous to regions of the  
 CC Ds elements which lie at the outermost ends of the transposons. A  
 CC series of low- and high-stringency PCR amplifications were  
 CC performed using the TAIL-PCR protocol. DNA fragments were produced  
 CC which corresponded to the genomic DNA that was directly adjacent to  
 CC the transposon insertion. Sequence analysis of PCR products from  
 CC tagged seedling lethal lines ET1158, GT6839 and ET5262 identified 3  
 CC novel genes (see AAF90548-50) each of which was essential for  
 CC Arabidopsis seedling growth and development. The essentiality of  
 CC the genes provides a means of discovering new herbicides. Screening  
 CC assays for identifying inhibitors that are potential herbicides are  
 CC provided. The invention is also applied to the development of  
 CC herbicide tolerant plants, and plant tissues, seeds and cells.

XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 3.1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

|||||

Db 2 tcgastwtsgwgtt 15

## RESULT 6

AAD06993

ID AAD06993 standard; DNA; 15 BP.

XX AAD06993;

XX 06-AUG-2001 (first entry)

XX PCR primer MDB285 to generate the flanking region of elite event MS-B2.

XX MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; ss.  
 OS Unidentified.  
 XX WO200131042-A2.  
 XX PD 03-MAY-2001.  
 XX PF 26-OCT-2000; 2000WO-EPI0680.  
 XX PR 29-OCT-1999; 99US-0430497.  
 XX PA (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX PI Weston B, De Beuckeleer M;  
 XX DR WPI; 2001-300517/31.  
 XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX  
 XX Example 3; Page 28; 53pp; English.  
 XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is degenerate PCR primer MDB285 used to generate  
 CC right (5') and left (3') border flanking region of elite event MS-B2.  
 XX  
 XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
 SQ

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 tcgastwtsgwgtt 15  
 Db | | | | | | | | | | | | | | | |  
 Db 2 tcgastwtsgwgtt 15

RESULT 7  
 AAF86681  
 ID AAF86681 standard; DNA; 15 BP.  
 XX  
 AC AAF86681;  
 XX  
 XX 25-JUL-2001 (first entry)  
 XX  
 DE Degenerate TAIL-PCR primer AD1, used for mapping insertion sites.  
 XX  
 KW Transgenic plant; cross-fertilisation; cross-breeding;  
 KW modified gene activity; expression; timing; tissue specificity;  
 KW modified nucleic acid structure; nucleic acid effector molecule;  
 KW activator; methylation enzyme; repressor; gyrase;  
 KW kinkase; topoisomerase; class I restriction enzyme; recognition sequence;  
 KW phenotype selection; Gramineae; Solanaceae; Leguminosae;  
 KW insertion site mapping; T-DNA; Ds element; rice; TAIL PCR primer;  
 KW thermal asymmetric interlaced PCR; ss.  
 XX Synthetic.  
 OS  
 XX WO200121781-A2.  
 PN  
 XX 29-MAR-2001.  
 PD  
 XX

PF 20-SEP-2000; 2000WO-US25778.  
 XX  
 PR 20-SEP-1999; 99US-0155036.  
 XX  
 XX (MOLE-) CENT APPL MOLECULAR BIOLOGY INT AGRIC.  
 XX  
 XX Jefferson RA, Killian A, Nottenburg C, Keese PK, Mayer J;  
 PI Stachel SE;  
 XX  
 XX WPI; 2001-335475/35.  
 DR  
 XX  
 XX Generating transgenic plants with modified genetic activity or nucleic  
 PT acid structure comprising crossing transgenic plant with gene for  
 PT nucleic acid effector with transgenic plant with binding site for  
 PT effector -  
 XX  
 XX Example 3; Page 27; 42pp; English.  
 XX The invention relates to a method of generating transgenic plants  
 CC that have modified gene activity or nucleic acid structure. The method  
 CC involves the cross-fertilisation of two transgenic plant lines to produce  
 CC seed, which is then grown to produce a plant. One of the parental plant  
 CC lines contains a gene encoding a non-native site-specific nucleic acid  
 CC effector molecule under the control of a minimal promoter and operably  
 CC linked to an endogenous enhancer element, and the other contains a  
 CC recognition sequence to which the nucleic acid effector molecule can  
 CC specifically bind. The nucleic acid effector molecule may be a  
 CC transcriptional activator, methylation enzyme, repressor, gyrase,  
 CC kinkase, topoisomerase, class I restriction enzyme, or single or double  
 CC stranded DNA or RNA. In another embodiment of the invention, three  
 CC transgenic plant lines are crossed, in which a transcriptional activator  
 CC gene, the minimal promoter and the recognition site are contained within  
 CC the three separate plant lines. In the transgenic cross-bred plant,  
 CC binding of the nucleic acid effector molecule to the recognition sequence  
 CC modifies gene expression, the timing of gene expression, the tissue  
 CC specificity of gene expression, or nucleic acid structure. Such plants  
 CC displaying a desired phenotype can then be selected. The method of the  
 CC invention can be used to generate transgenic plants, including Gramineae  
 CC (e.g., wheat, maize, rice, barley, rye, oats, sugarcane), Solanaceae  
 CC (e.g., potato, tomato, peppers), and Leguminosae (e.g., beans, soybean,  
 CC lentil, chickpea and peanut). Sequences AAF86681-AAF86683 represent  
 CC degenerate primers used in TAIL-PCR (thermal asymmetric interlaced PCR)  
 CC for physical mapping of T-DNA and Ds element insertion elements in  
 CC transgenic rice plants.  
 XX  
 XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
 SQ

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 tcgastwtsgwgtt 15  
 Db | | | | | | | | | | | | | | | |  
 Db 2 tcgastwtsgwgtt 15

RESULT 8  
 AAA89358  
 ID AAA89358 standard; DNA; 15 BP.  
 XX  
 AC AAA89358;  
 XX  
 XX 23-APR-2001 (first entry)  
 DT  
 XX Arabidopsis DNA primer AD1.  
 DE  
 XX Gene silencing; PCR primer; ss.  
 KW  
 XX Arabidopsis thaliana.  
 OS  
 XX WO200100801-A2.  
 PN  
 XX

PD 04-JAN-2001.  
XX  
PF 21-JUN-2000; 2000WO-EP05761.  
XX  
PR 23-JUN-1999; 99GB-0014623.  
XX  
PA (NOVS ) NOVARTIS AG.  
XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
XX  
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;  
XX  
DR WPI; 2001-137952/14.  
XX  
XX Novel gene encoding a protein that controls gene silencing, in  
PT particular silencing of plant genes -  
XX  
XX Example 4; Page 11; 48pp; English.  
XX  
XX This is the sequence of primer AD1, which is 1 of 7 degenerate  
CC primers (see AAA89358-84) expected to bind to Arabidopsis thaliana  
CC DNA flanking the site of a T-DNA insert. These primers were used  
CC with 3 specific, nested primers (see AAA89355-57), located close to  
CC the right border of T-DNA, in the cloning of the silencing gene  
CC (see AAA89353) of Arabidopsis thaliana. Genomic DNA from the plant  
CC containing only T-DNA co-segregating with a hygromycin resistant  
CC mutant phenotype was isolated. Gene silencing is useful as a  
CC molecular tool for regulating gene expression.  
XX  
XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
SQ

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
Db |||||  
2 tcgastwtsgwgtt 15

RESULT 9  
AAC86474  
ID AAC86474 standard; DNA; 15 BP.  
XX  
AC AAC86474;  
XX  
DT 01-MAR-2001 (first entry)  
XX  
DE Maize Mac2 gene degenerate PCR primer AD1.  
XX  
XX Tapetum-specific promoter; artificial male sterility; agriculture;  
KW crop yield; disease resistance; AMS; maize; MAC2; PCR primer; ss.  
XX  
XX Zea mays.  
OS  
XX  
XX WO200068403-A2.  
PN  
XX  
XX 16-NOV-2000.  
PD  
XX  
PF 10-MAY-2000; 2000WO-GB01789.  
XX  
XX 10-MAY-1999; 99GB-0010796.  
PR  
XX  
XX (BIOG-) BIOGEWA UK LTD.  
PA  
XX Paul W, Scott RJ, Hird D, Hodge R;  
PI  
XX WPI; 2001-016099/02.  
DR  
XX Novel nucleic acid sequences encoding tapetum specific promoters, pMAC2  
PT and pMAC20, or promoter sequences that control the expression of coding  
PT sequence homologous to the promoters, for artificial male sterility -  
XX

PS Example 2; Page 16; 45pp; English.  
XX  
XX The present invention provides tapetum-specific promoters which can be  
CC used to create artificial male sterility (AMS) systems in plants. These  
CC are useful in agriculture as they enable the production of plants with  
CC increased yield and greater disease resistance. They also result in  
CC uniform offspring, which have the same germination time, height of  
CC growth, susceptibility to disease, flowering time etc.  
XX  
XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
SQ

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
Db |||||  
2 tcgastwtsgwgtt 15

RESULT 10  
AAC64167/c  
ID AAC64167 standard; DNA; 30 BP.  
XX  
AC AAC64167;  
XX  
XX 21-FEB-2001 (first entry)  
DT  
XX  
XX Human pollinosis-associated gene 581 hybridisation probe, SEQ ID NO:8.  
DE  
XX  
XX Human; pollinosis-associated gene 581; IgE; immunoglobulin E;  
KW cedar pollen allergy; T-cell; reduced expression; detection;  
KW diagnosis; drug screening; allergic disease; expression analysis;  
KW hybridisation probe; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200065048-A1.  
PN  
XX 02-NOV-2000.  
PD  
XX 26-APR-2000; 2000WO-JP02732.  
PF  
XX 27-APR-1999; 99JP-0120492.  
PR  
XX (GENO-) GENOX RES INC.  
PA  
XX Nagasu T, Sugita Y, Kashiwabara T, Oshida T, Obayashi M, Gunji S;  
PI Obayashi I, Imai Y, Yoshida N, Ogawa K, Matsui K;  
XX  
XX WPI; 2000-687341/67.  
DR  
XX Pollenosis-associated gene 581 undergoing significantly low expression  
PT in subjects with high cedar pollen-specific IgE levels, useful in  
PT diagnosis of allergic diseases and screening drug candidates -  
XX  
XX Example 8; Page 24; 69pp; Japanese.  
PS  
XX The invention relates to the human pollinosis-associated gene 581 which  
CC exhibits significantly reduced expression in the T-cells of individuals  
CC with high cedar pollen-specific IgE (immunoglobulin E) levels. The gene  
CC was isolated from T-cells from individuals allergic to cedar pollen using  
CC the differential display method. The invention also relates also relates  
CC to the protein encoded by pollinosis-associated gene 581; to expression  
CC constructs and host cells comprising pollinosis-associated gene 581  
CC nucleic acids; pollinosis-associated gene 581 primers and probes;  
CC antibodies against the protein encoded by the gene; methods of detection  
CC of pollinosis-associated gene 581 nucleic acids; and a method of  
CC diagnosis of allergic diseases via the detection of pollinosis-associated  
CC gene 581 nucleic acids. The invention additionally encompasses methods of  
CC screening drug candidates for the treatment of allergic disease by  
CC measuring the expression of pollinosis-associated gene 581 in pollen

CC antigen-stimulated T-cells in the presence of a test compound relative  
 CC to a control. Pollinosis-associated gene 581 is useful in the diagnosis  
 CC of allergic diseases and in the screening of drug candidates for the  
 CC treatment of such diseases. The present sequence represents a  
 CC hybridisation probe used in human pollinosis-associated gene 581  
 CC expression analysis.

SQ Sequence 30 BP; 9 A; 9 C; 4 G; 8 T; 0 other;

Query Match 82.7%; Score 12.4; DB 21; Length 30;  
 Best Local Similarity 71.4%; Pred. No. 3.2e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 ||||:|:|:|:|  
 Db 30 TCGAGTATGGTGT 17

## RESULT 11

AAH21085  
 ID AAH21085 standard; DNA; 304 BP.  
 XX  
 AC AAH21085;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE Parallel detection of methylated genomic DNA associated DNA SEQ ID 31.  
 XX  
 KW DNA methylation; parallel detection; 5-unmethylated cytosine; CpG;  
 KW CpNG; amplification; transcription regulation; genetic imprinting;  
 KW tumorigenesis; primer; ss.  
 XX  
 OS Unidentified.

XX WO200142493-A2.

XX 14-JUN-2001.

XX 06-DEC-2000; 2000WO-DE04381.

XX 06-DEC-1999; 99DE-1059691.

XX (BPIG-) EPIGENOMICS AG.

XX Olek A, Plepenbrock C;

XX WPI; 2001-381705/40.

XX Parallel detection of the methylation pattern of many genomic DNA  
 XX regions, useful for detecting aberrant methylation, includes multiple  
 XX amplification of chemically modified DNA -

XX Disclosure; Page 63; 63pp; German.

XX This invention describes a novel method for the parallel detection of the  
 XX methylation status of genomic DNA (I) which involves a (I) sample being  
 XX treated chemically to convert 5-unmethylated cytosine to uracil,  
 XX thymidine or some other base having hybridization behavior different  
 XX that of C, then amplifying simultaneously at least 10 different fragments  
 XX (of fewer than 2 kb) using synthetic oligonucleotide (ON) primers. These  
 XX primers are based on regulatory, transcribed and/or translated segments  
 XX present in the sample after chemical treatment. The sequence context of  
 XX all, or some, of the CpG and CpNG motifs in the amplified products is  
 XX then determined. The method is used to detect aberrant methylation  
 XX patterns in the genome, these are implicated in regulation of  
 XX transcription, genetic imprinting and tumorigenesis. Many target regions  
 XX in the genome can be analyzed simultaneously and it is not essential to  
 XX know the sequence context of all targeted regions. Primers may be  
 XX designed for preferential amplification of particular segments of  
 XX interest (e.g. promoters and exons).

XX Sequence 304 BP; 37 A; 35 C; 131 G; 101 T; 0 other;

Query Match 82.7%; Score 12.4; DB 22; Length 304;  
 Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 ||||:|:|:|:|  
 Db 112 tcgagtttgaggt 125

## RESULT 12

AAC02398  
 ID AAC02398 standard; cDNA; 335 BP.

XX AC AAC02398;

XX XX

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 2396.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GEST ) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX P-PSDB; AAG02392.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 2396; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 XX mRNAs encoding secreted proteins. An ORF has been identified within the  
 XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 XX derived from 30 different tissues. EST sequences usually correspond  
 XX mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 XX well suited for isolating cDNA sequences derived from the 5' ends of  
 XX mRNAs and even in those cases where longer cDNA sequences have been  
 XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 XX mRNAs with intact 5' ends and can therefore be used to obtain full length  
 XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 XX gene therapy and chromosome mapping procedures. They are used to obtain  
 XX upstream regulatory sequences and to design expression and secretion  
 XX vectors.

XX Sequence 335 BP; 71 A; 75 C; 89 G; 89 T; 11 other;

Query Match 82.7%; Score 12.4; DB 21; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 ||||:|:|:|:|  
 Db 194 tcgagtttgaggt 207



```

RESULT 13
AAX51486
ID AAX51486 standard; cDNA; 336 BP.
XX
AC AAX51486;
XX
DT 21-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO:65.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906549-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01231.
XX
PR 01-AUG-1997; 97US-0905279.
XX
PA (GEST ) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153779/13.
XX
DR P-PSDB; AAY12708.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS Claim 1; Page 192; 522pp; English.
XX
CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12681 to
CC AAY12913, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 336 BP; 72 A; 75 C; 89 G; 89 T; 11 other;

Query Match 82.7%; Score 12.4; DB 20; Length 336;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
Db 195 tcgagtttgagtt 208
|||||:|:|:|

RESULT 14
AAX20974/c
ID AAX20974 standard; DNA; 388 BP.

```

```

XX AAX20974;
AC 05-MAY-1999 (first entry)
XX
DE Polynucleotide sequence from the genome of Treponema pallidum.
XX
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
XX
OS Treponema pallidum.
XX
PN WO9859034-A2.
XX
PD 30-DEC-1998.
XX
PF 23-JUN-1998; 98WO-US13041.
XX
PR 24-JUN-1997; 97US-0050667.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fraser CM;
XX
DR WPI; 1999-081273/07.
XX
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
XX
PS Claim 1; Page 998-999; 1150pp; English.
XX
CC AAX20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
SQ Sequence 388 BP; 121 A; 68 C; 77 G; 117 T; 5 other;

Query Match 82.7%; Score 12.4; DB 20; Length 388;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
Db 323 TCGAGTTTGGGT 310
|||||:|:|:|

RESULT 15
AAD06997
ID AAD06997 standard; DNA; 415 BP.
XX
AC AAD06997;
XX
DT 06-AUG-2001 (first entry)
XX
DE Right (5') border flanking region of elite event MS-B2.
XX
KW MS-B2 elite event; transgenic Brassica plant; transformation event;
KW male-sterility gene; ds.
XX
OS Chimeric - Agrobacterium sp.
OS Chimeric - Brassica sp.
XX
FH Key Location/Qualifiers
FT misc_feature 1..234
FT /tag= a
FT /note= "Corresponds to plant DNA"
FT misc_feature 235..415
FT /tag= b

```

FT /note= "Corresponds to T-DNA"

XX WO200131042-A2.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-EPI0680.

XX 29-OCT-1999; 99US-0430497.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX Weston B, De Beuckeleer M;

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -

PS Claim 11; Page 51; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is right (5') border flanking region of elite event  
 CC MS-B2.

XX SQ Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 82.7%; Score 12.4; DB 22; Length 415;  
 Best Local Similarity 71.4%; Pred. No. 3 6e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwtgtt 15

Db 2 tcgagtttggtgtt 15

Search completed: February 15, 2002, 19:01:14  
 Job time: 20718 sec

GenCore version 4.5  
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:07:24 ; Search time 353.79 Seconds  
 (without alignments)  
 9.602 Million cell updates/sec

Title: US-09-698-903B-4

Perfect score: 15  
 Sequence: 1 ntcgastwtsgwgtt 15

Scoring table: IDENTITY\_NUC  
 Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12.4	82.7	15	4	US-09-503-391-13
2	12.4	82.7	49272	1	US-08-614-770A-1
3	11.4	76.0	576	1	US-08-086-428B-1
4	11.4	76.0	576	2	US-08-468-570-1
5	11.4	76.0	576	2	US-08-290-665A-1
6	11.4	76.0	576	5	PCT-US95-10398-1
7	11.4	76.0	1268	4	US-09-046-894-30
8	11.4	76.0	2220	5	PCT-US95-13749-2
9	11.4	76.0	2341	3	US-09-187-049-11
10	11.4	76.0	2493	3	US-08-804-439A-11
11	11.4	76.0	2493	3	US-08-720-229-11
12	11.4	76.0	2652	2	US-08-953-492-1
13	11.4	76.0	2873	1	US-08-149-695-1
14	11.4	76.0	2873	1	US-08-377-228-1
15	11.4	76.0	3139	2	US-08-693-457-1
16	11.4	76.0	3139	4	US-09-265-731-1
17	11.4	76.0	12311	4	US-08-750-717-1
18	11.2	74.7	1567	1	US-08-181-271A-100
19	11.2	74.7	1567	1	US-08-449-315-100
20	11.2	74.7	1567	1	US-08-444-803-100
21	11.2	74.7	1567	1	US-08-449-043-100
22	11.2	74.7	1567	1	US-08-456-265A-100
23	11.2	74.7	1567	1	US-08-455-416-100
24	11.2	74.7	1567	1	US-08-455-244-100
25	11.2	74.7	1567	1	US-08-454-876-100
26	11.2	74.7	1567	2	US-08-457-364-100
27	11.2	74.7	1567	2	US-08-456-262-100

c 28	11.2	74.7	1567	2	US-08-456-240-100	Sequence 100, App
c 29	11.2	74.7	1567	2	US-08-455-736-100	Sequence 100, App
c 30	11.2	74.7	1567	2	US-08-971-217-100	Sequence 100, App
c 31	11.2	74.7	1567	4	US-09-350-600-100	Sequence 100, App
c 32	11.2	74.7	2068	1	US-08-309-341-1	Sequence 1, Appli
c 33	11.2	74.7	2068	1	US-08-608-267-1	Sequence 1, Appli
c 34	11.2	74.7	2068	1	US-08-608-452-1	Sequence 1, Appli
c 35	11.2	74.7	2068	1	US-08-608-224-1	Sequence 1, Appli
c 36	11.2	74.7	2068	2	US-08-967-149-1	Sequence 1, Appli
c 37	11.2	74.7	3117	1	US-08-172-331B-3	Sequence 3, Appli
c 38	11.2	74.7	3280	1	US-08-259-000-4	Sequence 4, Appli
c 39	11.2	74.7	3280	1	US-08-729-767-6	Sequence 6, Appli
c 40	11.2	74.7	4643	2	US-08-605-106-6	Sequence 6, Appli
c 41	11.2	74.7	7400	1	US-07-674-852-1	Sequence 1, Appli
c 42	11.2	74.7	7400	3	US-08-473-185-1	Sequence 1, Appli
c 43	11.2	74.7	7400	4	US-09-171-387-3	Sequence 3, Appli
c 44	11.2	74.7	12494	4	US-08-935-312-13	Sequence 13, Appli
c 45	11.2	74.7	12494	4	US-08-848-760B-33	Sequence 33, Appli

# ALIGNMENTS

RESULT 1  
 US-09-503-391-13  
 ; Sequence 13, Application US/09503391  
 ; Patent No. 6300091  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Patton, David A.  
 ; APPLICANT: Ashby, Carl S.  
 ; APPLICANT: Thomas, Carla R.  
 ; APPLICANT: McElver, John A.  
 ; APPLICANT: Budziszewski, Gregory J.  
 ; APPLICANT: Levin, Joshua Z.  
 ; TITLE OF INVENTION: Herbicide Target Genes and Methods  
 ; FILE REFERENCE: PB/5-30852A  
 ; CURRENT APPLICATION NUMBER: US/09/503,391  
 ; CURRENT FILING DATE: 2000-02-14  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 13  
 ; LENGTH: 15  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:  
 ; OTHER INFORMATION: oligonucleotide  
 US-09-503-391-13

Query Match 82.7%; Score 12.4; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
 Db 2 tcgastwtsgwgtt 15  
 |||||

RESULT 2  
 US-08-614-770A-1/c  
 ; Sequence 1, Application US/08614770A  
 ; Patent No. 5773267  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL  
 ; TITLE OF INVENTION: D29 SHUTTLE PHASMIDS AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
 ; STREET: 90 PARK AVENUE  
 ; CITY: NEW YORK  
 ; STATE: NEW YORK  
 ; COUNTRY: U.S.A.

ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE  
MEDIUM TYPE: DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/614,770A  
FILING DATE: MARCH 7, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ELIZABETH A. BOGOSIAN  
REGISTRATION NUMBER: 39,911  
REFERENCE/DOCKET NUMBER: 96700/402  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 49272  
TYPE: NUCLEIC ACID  
STRANDEDNESS: DOUBLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: OLIGONUCLEOTIDE  
DESCRIPTION: NO  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: MYCOBACTERIOPHAGE  
INDIVIDUAL ISOLATE: D29  
US-08-614-770A-1

Query Match 82.7%; Score 12.4; DB 1; Length 49272;  
Best Local Similarity 71.4%; Pred. No. 1e+02; 0; Indels 0; Gaps 0;  
Matches 10; Conservative 4; Mismatches 0;  
QY 2 tcgastwtsgwgtt 15  
Db 29087 TCGAGTTCGTGT 29074

RESULT 3  
US-08-086-428B-1  
; Sequence 1, Application US/08086428B  
; Patent No. 5514539  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R.H. AND  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 159  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/086,428B  
; FILING DATE: 29-JUN-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4070  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
ORIGINAL SOURCE:  
ORGANISM: homosapiens  
INDIVIDUAL ISOLATE: DK7  
US-08-086-428B-1

Query Match 76.0%; Score 11.4; DB 1; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgt 14  
Db 55 TCGAGTATCGTGT 67

RESULT 4  
US-08-468-570-1  
; Sequence 1, Application US/08468570  
; Patent No. 5871962  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 159  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,570  
; FILING DATE: 6-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/086,428  
; FILING DATE: 29-JUN-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4070US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 576 base pairs  
; TYPE: nucleic acid

;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; ORIGINAL SOURCE:  
;; ORGANISM: homosapiens  
;; INDIVIDUAL ISOLATE: DK7  
US-08-468-570-1

Query Match 76.0%; Score 11.4; DB 2; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14  
||||:|:|:|  
Db 55 TCGAGTATCGTGT 67

# RESULT 5

US-08-290-665A-1  
; Sequence 1, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:

;; APPLICANT: BUKH, J., MILLER, R.H. AND  
;; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
;; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
;; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
;; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
;; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
;; NUMBER OF SEQUENCES: 263  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: MORGAN & FINNEGAN  
;; STREET: 345 PARK AVENUE  
;; CITY: NEW YORK  
;; STATE: NEW YORK  
;; COUNTRY: USA  
;; ZIP: 10154

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: FLOPPY DISK  
;; COMPUTER: IBM PC COMPATIBLE  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: WORDPERFECT 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/290.665A  
;; FILING DATE: 15-AUG-1994  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: RICHARD W. BORK  
;; REGISTRATION NUMBER: 36,459  
;; REFERENCE/DOCKET NUMBER: 2026-4116  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 758-4800  
;; TELEFAX: (212) 751-8849  
;; TELEX: 421792

;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 576 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; ORIGINAL SOURCE:  
;; ORGANISM: homosapiens  
;; INDIVIDUAL ISOLATE: DK7  
US-08-290-665A-1

Query Match 76.0%; Score 11.4; DB 2; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14  
||||:|:|:|  
Db 55 TCGAGTATCGTGT 67

# RESULT 6

PCT-US95-10398-1  
; Sequence 1, Application PC/TUS9510398  
; GENERAL INFORMATION:

;; APPLICANT: BUKH, J., MILLER, R.H. AND  
;; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
;; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
;; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
;; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
;; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
;; NUMBER OF SEQUENCES: 263  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: MORGAN & FINNEGAN  
;; STREET: 345 PARK AVENUE  
;; CITY: NEW YORK  
;; STATE: NEW YORK  
;; COUNTRY: USA  
;; ZIP: 10154

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: FLOPPY DISK  
;; COMPUTER: IBM PC COMPATIBLE  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: WORDPERFECT 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US95/10398  
;; FILING DATE: 15-AUG-1995  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/086,428  
;; FILING DATE: 29 JUNE 1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/290/665  
;; FILING DATE: 15 AUGUST 1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: RICHARD W. BORK  
;; REGISTRATION NUMBER: 36,459  
;; REFERENCE/DOCKET NUMBER: 2026-4116  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 758-4800  
;; TELEFAX: (212) 751-6849  
;; TELEX: 421792

;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 576 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; ORIGINAL SOURCE:  
;; ORGANISM: homosapiens  
;; INDIVIDUAL ISOLATE: DK7  
PCT-US95-10398-1

Query Match 76.0%; Score 11.4; DB 5; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14  
||||:|:|:|  
Db 55 TCGAGTATCGTGT 67

# RESULT 7

US-09-046-894-30  
; Sequence 30, Application US/09046894  
; Patent No. 6190857  
; GENERAL INFORMATION:

;; APPLICANT: Ralph, David  
;; APPLICANT: An, Gang  
;; APPLICANT: O'Hara, Mark S.

```

; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: DIAGNOSIS OF DISEASE STATE USING mRNA
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,894
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,576
; FILING DATE: 24-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:014
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1268 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-046-894-30

Query Match 76.0%; Score 11.4; DB 4; Length 1268;
Best Local Similarity 69.2%; Pred. No. 3.5e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 3 cgastwtsgwgtt 15
Db 850 CGAGTTCGTGTT 862

RESULT 8
PCT-US95-13749-2/G
; Sequence 2, Application PC/TUS9513749
; GENERAL INFORMATION:
; APPLICANT: Amgen Inc.
; TITLE OF INVENTION: COMPOSITIONS FOR INCREASED
; BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13749
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 536..1024
; PCT-US95-13749-2

Query Match 76.0%; Score 11.4; DB 5; Length 2220;
Best Local Similarity 69.2%; Pred. No. 3.6e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwtg 14
Db 975 TCGAGTTGGAGT 963

RESULT 9
US-09-187-049-11/c
; Sequence 11, Application US/09187049
; Patent No. 6117666
; GENERAL INFORMATION:
; APPLICANT: Lamppa, Gayle K.
; TITLE OF INVENTION: PLASTID PROTEOLYTIC PROCESSING ENZYME
; TITLE OF INVENTION: THAT CLEAVES PRECURSOR POLYPEPTIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BRINKS HOFER GILSON & LIONE
; STREET: P.O. Box 10395
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/187,049
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/695,177
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Martin, Alice O.
; REGISTRATION NUMBER: 35,601
; REFERENCE/DOCKET NUMBER: 7814/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 321-4200
; TELEFAX: 312 321-4299
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2341 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis CPE
; US-09-187-049-11

Query Match 76.0%; Score 11.4; DB 3; Length 2341;
Best Local Similarity 69.2%; Pred. No. 3.6e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 tcgastwtsgwt 14  
 ||||:|:|:|  
 Db 599 TCGACTATGGAGT 587

## RESULT 10

US-08-804-439A-11  
 ; Sequence 11, Application US/08804439A  
 ; Patent No. 6015565  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, Timothy M.  
 ; APPLICANT: Bosch, Marnix L.  
 ; APPLICANT: Strand, Kurt  
 ; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV  
 ; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
 ; NUMBER OF SEQUENCES: 113  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Ste 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/804,439A  
 ; FILING DATE: February 21, 1997  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haile, Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 09176/004001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 678-5070  
 ; TELEFAX: (619) 678-5099  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2493 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-804-439A-11

Query Match 76.0%; Score 11.4; DB 3; Length 2493;  
 Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwt 14  
 ||||:|:|:|  
 Db 1369 TCGAGTATCGTGT 1381

## RESULT 11

US-08-720-229-11  
 ; Sequence 11, Application US/08720229  
 ; Patent No. 6022542  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, Timothy M.  
 ; APPLICANT: Bosch, Marnix L.  
 ; APPLICANT: Strand, Kurt  
 ; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV  
 ; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
 ; NUMBER OF SEQUENCES: 100  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morrison & Foerster  
 ; STREET: 755 Page Mill Road

; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304-1018  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/720,229  
 ; FILING DATE: 26-SEP-1996  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Schliff, J. Michael  
 ; REGISTRATION NUMBER: 40,253  
 ; REFERENCE/DOCKET NUMBER: 29938-20002.00  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 813-5600  
 ; TELEFAX: (415) 494-0792  
 ; TELEX: 706141  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2493 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-720-229-11

Query Match 76.0%; Score 11.4; DB 3; Length 2493;  
 Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwt 14  
 ||||:|:|:|

Db 1369 TCGAGTATCGTGT 1381

## RESULT 12

US-08-953-492-1  
 ; Sequence 1, Application US/08953492  
 ; Patent No. 5849555  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brown, James  
 ; APPLICANT: Jaworski, Deborah  
 ; APPLICANT: Lawlor, Elizabeth  
 ; APPLICANT: Wang, Min  
 ; TITLE OF INVENTION: NOVEL vals  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406-0939  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/953,492  
 ; FILING DATE: 17-OCT-1997  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/844,064  
 ; FILING DATE: 18-APR-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 9607991.8  
 ; FILING DATE: 18-APR-1996  
 ; ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31458-4/1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2652 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-953-492-1

Query Match 76.0%; Score 11.4; DB 2; Length 2652;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tgcagtttgaggt 14  
Db 1892 tgcagtttgaggt 1904

RESULT 13  
US-08-149-695-1  
; Sequence 1, Application US/08149695  
; Patent No. 5412085  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Rebecca L.  
; APPLICANT: Lonsdale, David M.  
; TITLE OF INVENTION: A Pollen-Specific Promoter From Maize  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McAndrews, Held and Malloy  
; STREET: 500 W. Madison, 34th Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60661

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/149,695  
FILING DATE:

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/911,532  
FILING DATE: 09-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pochopien, Donald J.  
REGISTRATION NUMBER: 32167  
REFERENCE/DOCKET NUMBER: 92 P139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312)707-8889  
TELEFAX: (312)707-9155

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Zea mays  
STRAIN: Line W22

TISSUE TYPE: Pollen  
US-08-149-695-1

Query Match 76.0%; Score 11.4; DB 1; Length 2873;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 3 cgastwtsgwgtt 15  
Db 111 CGAGTTTGGAGTT 123

RESULT 14  
US-08-377-228-1  
; Sequence 1, Application US/08377228  
; Patent No. 5545546  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Rebecca L.  
; APPLICANT: LONSDALE, David M.  
; TITLE OF INVENTION: A Pollen-Specific Promoter From Maize  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/377,228  
FILING DATE: 24-JAN-1995  
CLASSIFICATION: 536

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/149,695  
FILING DATE: 09-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/911,532  
FILING DATE: 09-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 33229/290/PIHI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-377-228-1

Query Match 76.0%; Score 11.4; DB 1; Length 2873;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 3 cgastwtsgwgtt 15  
Db 111 CGAGTTTGGAGTT 123

RESULT 15  
US-08-693-457-1  
; Sequence 1, Application US/08693457



; Patent No. 5880330  
; GENERAL INFORMATION:  
; APPLICANT: Weigel et al., Detlef  
; TITLE OF INVENTION: SHOOT MERISTEM SPECIFIC PROMOTER  
; TITLE OF INVENTION: SEQUENCES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/693,457  
; FILING DATE: 07-AUG-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07251/012001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3139 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 571..1900  
; US-08-693-457-1

Query Match 76.08; Score 11.4; DB 2; Length 3139;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 cgastwtsgwgtt 15  
|||:|:|:|:|  
Db 1546 CGACITGGAGTT 1558

Search completed: February 15, 2002, 19:07:28  
Job time: 14622 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:53:51 ; Search time 2553.1 Seconds  
(without alignments)  
148.617 Million cell updates/sec

Title: US-09-698-903B-11

Perfect score: 23  
Sequence: 1 gaaatccatgtaagcagcagg 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hgt: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rod: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	DB	ID	Description
1	23	100.0	23	6	AX127758	Sequence AX127758
2	23	100.0	415	6	AX127755	Sequence AX127755
3	23	100.0	416	6	AX127757	Sequence AX127757
c 4	18.8	81.7	41760	2	AC090679	Homo sapi AC090679
c 5	18.8	81.7	89350	8	ATT12H17	Arabidops AT12H17
c 6	18.8	81.7	156515	2	AC022947	Homo sapi AC022947
c 7	18.8	81.7	195577	8	ATCHRIV57	Arabidops ATCHRIV57
c 8	18.8	81.7	208684	2	AC072048	Mus muscu AC072048
c 9	18.4	80.0	168524	2	AC046148	Mus muscu AC046148
c 10	18.4	80.0	172650	2	AC040920	Homo sapi AC040920
c 11	18.2	79.1	535	5	AF118429	Falco per AF118429
c 12	18.2	79.1	146285	9	AC005083	Homo sapi AC005083
c 13	18.2	79.1	183634	2	AC023167	Mus muscu AC023167
c 14	18.2	79.1	228809	2	AC021443	Homo sapi AC021443
c 15	17.8	77.4	23486	8	AC007294	Arabidops AC007294
c 16	17.8	77.4	26275	3	U29377	Caenorhabdi U29377
c 17	17.8	77.4	94835	9	HS0316D7	Human DNA HS0316D7
c 18	17.8	77.4	139959	33	AC013597	Homo sapi AC013597
c 19	17.8	77.4	146746	9	HS03122	Human DNA s HS03122
c 20	17.8	77.4	149755	2	AC092795	Homo sapi AC092795
c 21	17.8	77.4	168749	2	AC008573	Homo sapi AC008573
c 22	17.8	77.4	177703	9	AC019046	Homo sapi AC019046
c 23	17.8	77.4	179665	2	AC010246	Homo sapi AC010246
c 24	17.8	77.4	188537	2	AC090071	Homo sapi AC090071
c 25	17.8	77.4	213890	2	AC013461	Homo sapi AC013461
c 26	17.8	77.4	253038	2	AC008930	Homo sapi AC008930
c 27	17.4	75.7	25042	9	HS380A1	Human DNA s HS380A1
c 28	17.4	75.7	89939	2	AC022667	Homo sapi AC022667
c 29	17.4	75.7	106392	9	AC007316	Homo sapi AC007316
c 30	17.4	75.7	113033	9	AC005070	Homo sapi AC005070
c 31	17.4	75.7	118904	9	HS238G2	Human DNA HS238G2
c 32	17.4	75.7	134506	9	HS357I16	Homo sapi HS357I16
c 33	17.4	75.7	136497	2	HS0448H8	Homo sapi HS0448H8
c 34	17.4	75.7	155316	2	AC078902	Homo sapi AC078902
c 35	17.4	75.7	155932	1	CJ11168X6	Campyloba CJ11168X6
c 36	17.4	75.7	156014	5	AP003796	Gallus ga AP003796
c 37	17.4	75.7	171368	9	HSU95738	Human chrom HSU95738
c 38	17.4	75.7	171868	9	AL359697	Human DNA AL359697
c 39	17.4	75.7	194873	2	AL592440	Homo sapi AL592440
c 40	17.4	75.7	256073	9	AE006464	Homo sapi AE006464
c 41	17.2	74.8	466	11	HSPE53B06	R. catesbeia HSPE53B06
c 42	17.2	74.8	522	5	RCABCRTST	X87114 R. catesbeia X87114
c 43	17.2	74.8	985	3	AB002187	Epilachna AB002187
c 44	17.2	74.8	985	3	AB002188	Epilachna AB002188
c 45	17.2	74.8	985	3	AB002189	Epilachna AB002189

ALIGNMENTS

RESULT 1

AX127758	Sequence 11 from Patent WO0131042.	23 bp	DNA	PAT	15-MAY-2001
LOCUS	AX127758				
DEFINITION	AX127758				
ACCESSION	AX127758.1	GI:14134405			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					

synthetic construct.  
synthetic construct.  
artificial sequence.  
1 (bases 1 to 23)  
Weston,B. and de Beuckeleer,M.  
Male-sterile brassica plants and methods for producing same  
Patent: WO 0131042-A 11 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
Location/Qualifiers  
1. .23  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer MDB371"

```

BASE COUNT      9 a      4 c      7 g      3 t
ORIGIN

Query Match      100.0%; Score 23; DB 6; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgagcagcagg 23
|||||
Db 1 GAAATCCATGTTAAAGCAGCAGG 23

RESULT 2
AX127755
LOCUS AX127755 415 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 8 from Patent WO0131042.
ACCESSION AX127755
VERSION AX127755.1 GI:14134402
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 415)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 8 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source
1. 415
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="5' border flanking region of elite event MS-B2"
misc_feature 1..234
/note="plant DNA"
misc_feature 235..415
/note="T-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN

Query Match      100.0%; Score 23; DB 6; Length 415;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgagcagcagg 23
|||||
Db 152 GAAATCCATGTTAAAGCAGCAGG 174

RESULT 3
AX127757/c
LOCUS AX127757 416 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 10 from Patent WO0131042.
ACCESSION AX127757
VERSION AX127757.1 GI:14134404
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 416)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source
1. 416
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="3' border flanking region of elite event MS-B2"
misc_feature 1..193
misc_feature 194..416

```

```

BASE COUNT      137 a      72 c      54 g      152 t      1 others
ORIGIN

Query Match      100.0%; Score 23; DB 6; Length 416;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgagcagcagg 23
|||||
Db 263 GAAATCCATGTTAAAGCAGCAGG 241

RESULT 4
AC090679/c
LOCUS AC090679 41760 bp DNA HTG 27-JUN-2001
DEFINITION Homo sapiens chromosome 12q clone RP11-8003, *** SEQUENCING IN
PROGRESS ***; 14 unordered pieces.
ACCESSION AC090679
VERSION AC090679.3 GI:14190581
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 41760)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbieria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Cartoll,L., Dederich,D.A., Delaney,K.J., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escottio,M., Falls,T., Ferraguto,D., Flegg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Joudah,S., Karlsson,B., Jia,Y., Johnson,R., Jolivet,S.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogun,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 41760)
REFERENCE
Worley,K.C.
Direct Submission
Submitted (08-MAR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

COMMENT

Baylor Plaza, Houston, TX 77030, USA  
On May 23, 2001 this sequence version replaced gi:13470121.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HCQK  
Center clone name: RP11-8003  
----- Summary Statistics  
Sequencing vector: M13: L08821  
Chemistry: Dye-terminator Big Dye: 50% of reads  
Chemistry: Dye-terminator Big Dye: 50% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 91506 bases at least Q40  
Consensus quality: 101714 bases at least Q30  
Consensus quality: 106393 bases at least Q20  
Estimated insert size: 81455; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-gel estimation  
Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 6354: contig of 6354 bp in length  
\* 6355: gap of unknown length  
\* 6455: contig of 4439 bp in length  
\* 10893: gap of unknown length  
\* 10894: contig of 2562 bp in length  
\* 10994: gap of unknown length  
\* 13555: contig of 3979 bp in length  
\* 13656: gap of unknown length  
\* 17634: contig of 3029 bp in length  
\* 17735: gap of unknown length  
\* 20764: contig of 2666 bp in length  
\* 20864: gap of unknown length  
\* 23530: contig of 2193 bp in length  
\* 23630: gap of unknown length  
\* 25823: contig of 2109 bp in length  
\* 25923: gap of unknown length  
\* 28032: contig of 2408 bp in length  
\* 28132: gap of unknown length  
\* 30340: contig of 2059 bp in length  
\* 30640: gap of unknown length  
\* 32699: contig of 2339 bp in length  
\* 32799: gap of unknown length  
\* 35138: contig of 2058 bp in length  
\* 35238: gap of unknown length  
\* 37296: contig of 2209 bp in length  
\* 37396: gap of unknown length  
\* 39605: contig of 2056 bp in length.  
\* 39705: contig of 2056 bp in length.

FEATURES  
source

1. .41760  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12q"  
/clone="RP11-8003"

BASE COUNT 12256 a 7423 c 7257 g 13504 t 1320 others  
ORIGIN

Query Match 81.7%; Score 18.8; DB 2; Length 41760;  
Best Local Similarity 90.9%; Pred. No. 88;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaaatccatgtaagcagcagg 22  
||| ||||| ||||| |||||  
Db 19612 GAATTCATGTAAGTCAGG 19591

RESULT 5  
ATT12H17/c

LOCUS 89350 bp DNA PLN 03-FEB-1998  
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T12H17 (ESSAI  
project).  
ACCESSION AL021635  
VERSION AL021635.1 GI:2827538  
KEYWORDS  
SOURCE thale cress.

ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 89350)  
AUTHORS Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,  
Duesterhoeft,A., Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 89350)  
AUTHORS EU Arabidopsis sequencing project.  
TITLE Direct Submission  
JOURNAL Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
E-mail: michael.bevan@bsrc.ac.uk

FEATURES  
source

1. .89350  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
669. .1151  
/chromosome="4"  
669. .1151  
/gene="T12H17.10"  
669. .1151  
/gene="T12H17.10"  
/number=1  
669. .1151  
/gene="T12H17.10"  
/note="similarity to auxin-induced protein 10A, Glycine  
max., PIR2:J01099"  
/codon\_start=1  
/product="predicted protein"  
/protein\_id="CAAL6547.1"  
/db\_xref="GI:2827539"  
/db\_xref="SPTREMBL:O49643"  
/translation="MKRLRGFKIGHRFVKIFKWIIRSRNQTGKROCLTGILNPVTKI  
YSLARCLRGANFLCGMKPGOTRLGNEPKTPVPRGHLVHVHVGESGDTRRVVPV  
IYNHPLFGLLEQAERYVEQPGRIWPCRVSDFEKVQMRIAADHCRKSTFKIL  
"  
2406. .2754  
/gene="T12H17.20"  
/number=1  
2406. .2983  
/gene="T12H17.20"  
join(2406. .2754,2844. .2983)  
/gene="T12H17.20"  
/note="similarity to predicted proteins, Arabidopsis  
thaliana"  
/codon\_start=1  
/product="predicted protein"  
/protein\_id="CAAL6548.1"  
/db\_xref="GI:2827540"  
/db\_xref="SPTREMBL:O49644"  
/translation="MAYTNKVAVAVGAVVFLAVVMNPRTEAOTYKLDRLCYVMIP  
DILEECFTHRLRXPTECCNDLNKATMTQYDCLDNFSLSFSDLSRTSAGVLLKRC  
DVSHKYMCAAKNKGAGGRNSTTCDNSITNTSVGGKNKRVATSMASFAFLVLLFV  
MF"  
2755. .2843  
intron

```

/gene="T12H17.20"
/number=1
2844..2983
/gene="T12H17.20"
/number=2
3327..3400
/note="Asn-trna_(GTT)"
4696..5489
/gene="T12H17.30"
join(4696..5035,5407..5489)
/gene="T12H17.30"
/codon_start=1
/product="putative protein"
/protein_id="CAA16549.1"
/db_xref="GI:2827541"
/db_xref="SPTREMBL:O49645"
/translation="MAYTNKVTISAAVATMMLFLAVTIVDAQSMPPMPKFNVPVCALAD
LPNIVQICIFNLDLTPEECNDLKSSSTIQVNCCLDNFIAHPSNGNISQARYDLVNS
ACGVADKFAKGGDASGGSTNKIAASWVLLGLVASLFF"
4696..5035
/gene="T12H17.30"
/number=1
5036..5406
/gene="T12H17.30"
/number=1
5407..5489
/gene="T12H17.30"
/number=2
join(5695..6940,7035..7105)
/gene="T12H17.40"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAA16550.1"
/db_xref="GI:2827542"
/db_xref="SPTREMBL:O49646"
/translation="NAVAVTAVLEFLAVIAPOWTEKKPPRPDSDTSGTSGRRDRT
MGPLSTIGVQNCYATLNAFSPKCECKDLKASKREVTCCLNNVIAHPDPLYTNTNOV
NKACGVLDKACDVAPFTTIVTYNYDMIIIAVIQDP"
6395..6940
/gene="T12H17.40"
/number=1
6955..7105
/gene="T12H17.40"
6941..7034
/gene="T12H17.40"
/number=1
7035..7105
/gene="T12H17.40"
/number=2
complement(8019..9209)
/gene="T12H17.50"
/number=1
8019..9209
/gene="T12H17.50"
complement(8019..9209)
/gene="T12H17.50"
complement(8019..9209)
/gene="T12H17.50"
/note="similarity to predicted protein, Arabidopsis
thaliana"
/codon_start=1
/product="predicted protein"
/protein_id="CAA16551.1"
/db_xref="GI:2827543"
/db_xref="SPTREMBL:O49647"
/translation="MEKNQNPNTWSDLPDLLNLVFKRLSFANPROAKSVCSWYSAS
KQSVPKNQIPWLMFLPKDKNNKNSCTIFNPEDKQLYQTQDLGYEPAKSVCLAYI
GSWLMDQSKYNLILNPTVEKIGLPAISQQQGVKVDQTIIDDFLTDDHNNHVL
FKGNNTVTPVFWIDEKTEYIALWGLGVCVYAKNGDKLWNOIPEIILDSLDWYK
DKLISYSYRNLFILDFSGEIPKFAQCQFMHYVSEMLSPRQLSNWCVAETKLV
VTVGVDLMLRHSRQSFNVKFFSYSGTFLDKYELADSLGDEAMLDLIGITVL
ANEVEGLHRTNIYFSDSHDTTTKDLFLFNRETREMEPLHKFDFDCSLLELSARWFLPS

```

FSHT"

9990..113709

/gene="T12H17.60"

join(9990..10023,10239..10535,10564..10664,10910..11292,

11755..11809,11898..12053,12143..12311,12391..12455,

12534..12674,12769..12885,13024..13096,13244..13380,

13468..13525,13612..13709)

/gene="T12H17.60"

/note="similarity to p48, Homo sapiens, PATCH:G904032

contains EST gb:Aa605571, H76919, H76638, R30271, Z25650"

/codon\_start=1

/product="HSP associated protein like"

/protein\_id="CAA16552.1"

/db\_xref="GI:2827544"

/db\_xref="SPTREMBL:O49648"

/translation="MRRTWLAEESSVAIAPLLARPOSTMRPMTPECATVMPDLLEK

CFATGSVPTEDCTDLKATSTQVTCCLDNVIANPAVSNITGYSKAITTKCGVFDK

YSCDTSKVISNKGAFVINGFWCVYKKEKKEEAAYAMGVLTLLGLNLIVAH

GTRQONPVCKLILFQHFQFCISIRFSANKAKRSKTDGAYPFRNEKVLQKPY

LPFQILIEIETRLATKTRLLRLRPPTLFLSLSRSEADFSQKDFNEAYLGAK

IPTGVHEEDKTPKRSVVVEESDDMDDETEYVKPVEEEDEIDEVSDVELEGTVE

PNDPPKMGDSVEVTDENREAQAEAKKALSEGNDFAIHLRAITLNPSTA

IMYGRASVYIKLKPNAIRDAALAEINPDASAKYKSRGMARMLGEWAPAAKDLH

LASTIDYDEEISAVLKVPEPNAHKLSEHRRKYDRLKREDEKKAERDLRRRAEAOAA

YDKAKKEQSSSRSPSGGFGGMPGCGGMPAGMGCGMPAGMGCGMPGCGGMP

CGMDFSKILNDPELMTATSDPEVMAALQDVMMKPNANLAKHOANPKVAPVIAKMMGKFA

GPQ"

9990..10023

/gene="T12H17.60"

/number=1

10024..10238

/gene="T12H17.60"

/number=1

10239..10535

/gene="T12H17.60"

/number=2

10536..10563

/gene="T12H17.60"

/number=2

10564..10664

/gene="T12H17.60"

/number=3

10665..10909

/gene="T12H17.60"

/number=3

10910..11292

/gene="T12H17.60"

/number=4

11293..11754

/gene="T12H17.60"

/number=4

11755..11809

/gene="T12H17.60"

/number=5

11810..11897

/gene="T12H17.60"

/number=5

11898..12053

/gene="T12H17.60"

/number=6

12054..12142

Query Match 81.7%; Score 18.8; DB 8; Length 89350;

Best Local Similarity 90.9%; Pred. No. 95;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgaggg 23

|||||

Db 15674 AAATCCATGTAAGCAGTAGG 15653

RESULT 6

AC022947/c

```

LOCUS       AC022947       156515 bp       DNA             HTG             12-MAR-2000
DEFINITION   Homo sapiens clone RP11-8003, WORKING DRAFT SEQUENCE, 13 unordered
              pieces.
ACCESSION    AC022947
VERSION      AC022947.2  GI:7229879
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS      1 (bases 1 to 156515)
              Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens, clone RP11-8003
REFERENCE    2 (bases 1 to 156515)
              Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
              Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
              Choepeil,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
              DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenebor,J.,
              Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
              Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
              Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
              Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
              McPheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
              Norman,C.H., O'Connor,T., O'Donnell,P., Olivares,T.M., Peterson,K.,
              Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
              Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
              Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
              Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
              Zimmer,A. and Zody,M.
DIRECT SUBMISSION
              Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              On Mar 12, 2000 this sequence version replaced gi:6921807.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              ----- Project Information
              Center project name: L6397
              Center clone name: 80_O_3
              ----- Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 147422 bases at least Q40
              Consensus quality: 151928 bases at least Q30
              Consensus quality: 153939 bases at least Q20
              Insert size: 150000; agarose-fp
              Insert size: 155315; sum-of-contents
              Quality coverage: 4.8 in Q20 bases; agarose-fp
              Quality coverage: 4.6 in Q20 bases; sum-of-contents
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 13 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 1026: contig of 1026 bp in length
              * 1027 1126: gap of 100 bp
              * 1127 5787: contig of 4661 bp in length
              * 5788 5887: gap of 100 bp
              * 5888 11595: contig of 5708 bp in length
              * 11596 11695: gap of 100 bp

```

```

* 11696 14431: contig of 2736 bp in length
* 14432 14531: gap of 100 bp
* 14532 20088: contig of 5557 bp in length
* 20089 20188: gap of 100 bp
* 20189 25619: contig of 9431 bp in length
* 25620 29719: gap of 100 bp
* 29720 38087: contig of 8368 bp in length
* 38088 38187: gap of 100 bp
* 38188 50251: contig of 12064 bp in length
* 50252 50351: gap of 100 bp
* 50352 60663: contig of 10312 bp in length
* 60664 60763: gap of 100 bp
* 60764 75537: contig of 14774 bp in length
* 75538 75637: gap of 100 bp
* 75638 93139: contig of 17502 bp in length
* 93140 93239: gap of 100 bp
* 93240 114372: contig of 21133 bp in length
* 114373 11472: gap of 100 bp
* 11473 156515: contig of 42043 bp in length.
FEATURES             Location/Qualifiers
     source            1..156515
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="RP11-8003"
                        /clone_lib="RPCI-11 Human Male BAC"
     misc_feature       1..1026
                        /note="assembly_fragment"
     misc_feature       1127..5787
                        /note="assembly_fragment"
     misc_feature       5888..11595
                        /note="assembly_fragment"
     misc_feature       11696..14431
                        /note="assembly_fragment"
                        /note="clone_end:SP6"
                        /vector_side:left
     misc_feature       14532..20088
                        /note="assembly_fragment"
     misc_feature       20189..29619
                        /note="assembly_fragment"
                        /note="clone_end:T7"
                        /vector_side:left
     misc_feature       29720..38087
                        /note="assembly_fragment"
     misc_feature       38188..50251
                        /note="assembly_fragment"
     misc_feature       50352..60663
                        /note="assembly_fragment"
     misc_feature       60764..75537
                        /note="assembly_fragment"
     misc_feature       75638..93139
                        /note="assembly_fragment"
     misc_feature       93240..114372
                        /note="assembly_fragment"
     misc_feature       114473..156515
                        /note="assembly_fragment"
BASE COUNT  49638 a 27072 g 27170 t 1205 others
ORIGIN
Query Match      81.7%; Score 18.8; DB 2; Length 156515;
Best Local Similarity 90.9%; Pred. No.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gaaatccatgtaaacgacgag 22
   ||| ||||| ||||| |||||
Db 76912 GAATTCATGTAAGTCAGG 76891
RESULT 7
ATCHRIV57/c
LOCUS      ATCHRIV57 199577 bp      DNA             16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 57.
ACCESSION  AL161557

```

VERSION	AL161557.2	GI:7269071	
KEYWORDS	thale cress.		
SOURCE	Arabidopsis thaliana		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 64578 to 155510)		
REFERENCE	Wedler,H., Wambutt,R., Mewes,H.W., Lemcke,K. and Mayer,K.F.X. Unpublished		
JOURNAL	2 (bases 141638 to 141797)		
REFERENCE	Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W., Lemcke,K. and Mayer,K.F.X. Unpublished		
JOURNAL	3 (bases 143670 to 199577)		
REFERENCE	Hilbert,H., Braun,M., Holzer,E., Brandt,A., Dueterhoeft,A., Mewes,H.W., Lemcke,K. and Mayer,K.F.X. Unpublished		
JOURNAL	4 (bases 48909 to 64829)		
REFERENCE	Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and Mayer,K.F.X. Unpublished		
JOURNAL	5 (bases 1 to 199577)		
REFERENCE	EU Arabidopsis sequencing, project.		
AUTHORS	Direct Submission		
TITLE	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de		
JOURNAL	Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk		
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> this fragment has an overlap with ATCHRIV56 at the 5' end and an overlap with ATCHRIV58 at the 3' end.		
FEATURES	source	location/Qualifiers	
gene		1..199577	/organism="Arabidopsis thaliana"
exon		/variety="Columbia"	
		/db_xref="taxon:3702"	
		/chromosome="4"	
		6674..9057	
		/gene="AT4g22260"	
		complement(6674..6882)	
		/gene="AT4g22260"	
		/number=1	
gene		complement(join(6674..6882,7268..7339,7507..7575,7667..7775,7863..7978,8361..8454,8588..8777,8909..9057))	
		/gene="AT4g22260"	
		complement(join(6674..6882,7268..7339,7507..7575,7667..7775,7863..7978,8361..8454,8588..8777,8909..9057))	
		/gene="AT4g22260"	
CDS		/note="similarity to alternative oxidase, Mangifera indica, PIR2:545035"	
		contains EST gb:142793, AI994896.1, AI997645.1, R30022, AA395166	
		/codon_start=1	
		/product="putative protein"	
		/protein_id="CAB79181.1"	
		/db_xref="GI:7269072"	
		/translation="MAALISGSSTLTISRPLVLRRAAVSYSSSHRLHLPLSS RRLLRNHRVQATILQDDEKVVVEESFKASTGTGTEPLEPNSSSTSAFTWII KLEQGVNFTDSDIKITDLYRDRYARFVLEIARVPVFAFMSVLMHYETGWR RADLYKVFADSNEMHLLIMEELGGNSWDFDLAQHIAFYEVMTVFYILSPRM AYHPSCEVHAYETDKFLKASGEELKNMADPAIVKYITGGDLILFENLYDFVNI RDDAEHCKTHRAQTIGLSLRSPSILEDDEESGCVPEAHCEGIVDCLAKKSIT S"	
intron		complement(6883..7267)	
exon		/number=1	
		complement(7268..7339)	
		/gene="AT4g22260"	
		/number=2	
intron		complement(7340..7506)	
		/number=2	
exon		complement(7507..7575)	
		/gene="AT4g22260"	
		/number=3	
intron		complement(7576..7666)	
exon		complement(7667..7775)	
		/gene="AT4g22260"	
		/number=4	
intron		complement(7776..7862)	
		/number=4	
exon		complement(7863..7978)	
		/gene="AT4g22260"	
		/number=5	
intron		complement(7979..8360)	
		/number=5	
exon		complement(8361..8454)	
		/gene="AT4g22260"	
		/number=6	
intron		complement(8455..8587)	
		/number=6	
exon		complement(8588..8777)	
		/gene="AT4g22260"	
		/number=7	
intron		complement(8778..8908)	
		/number=7	
exon		complement(8909..9057)	
		/gene="AT4g22260"	
		/number=8	
exon		10103..10244	
		/gene="AT4g22270"	
		/number=1	
gene		10103..12393	
		/gene="AT4g22270"	
		join(10103..10244,10395..10516,11327..11779,11868..12174,12254..12393)	
CDS		/note="contains EST gb:AI997575.1"	
		/codon_start=1	
		/product="hypothetical protein"	
		/protein_id="CAB79182.1"	
		/db_xref="GI:7269073"	
		/translation="NATNGSSSEHLPEPTRSAIINKQOPEPPASRFTMSLVLFWDQ SNEGFSVLSIWSKFCMRREFLDKLDVSDKVRIRYEAEIQSLKLMIFVLPSSL TLEAIRWNIWISGNFIPIIINPILSHVACTQLQSSWLYNSLFIICILYKITCH LQTLRLDDFAFCFASIEIDVRSALGEHQIRNRIRIVSHRRFILLILVLTATQFM ALLTTRASVAVNIYEVGELALCSLVTGVICLSRSATKITHKAQSVTSLAAKNVVC ATVDSDHLDGETPTGSIIESQVSLRGNAIETSDDEGEODDLDNTKIHPIYANTIS YKQRAVLYLENKAGITVYGFVDRSWLNTIFGIELALLLWLLNKITGILA"	
		10245..10394	
		/gene="AT4g22270"	
		/number=1	
		10395..10516	
		/gene="AT4g22270"	
		/number=2	
intron		10517..11326	
		/gene="AT4g22270"	
		/number=2	
trna		11168..11253	
		/gene="AT4g22270"	
		/note="Tyr-tRNA (GTA)"	
exon		11327..11779	
		/gene="AT4g22270"	
		/number=3	
intron		11780..11867	
		/gene="AT4g22270"	
		/number=3	
exon		11868..12174	
		/gene="AT4g22270"	
		/number=4	
intron		12175..12253	
		/gene="AT4g22270"	



/number=4  
/gene="AT4g22270"  
/number=5  
12761..15111  
/gene="AT4g22280"  
complement(12761..13015)  
/gene="AT4g22280"  
/number=1  
complement(join(12761..13015,13081..13227,13379..14157,14209..14314,14995..15111))  
/gene="AT4g22280"  
complement(join(12761..13015,13081..13227,13379..14157,14209..14314,14995..15111))  
/gene="AT4g22280"  
/note="similarity to hypothetical protein, chr.4, Arabidopsis thaliana, PATCHK:G2392763 contains EST gb:A1998861.1, A1993240.1"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB79183.1"  
/db\_xref="GI:7269074"  
/translation="MEKQVMTDETCTVGNPATIVNVNGLNISHOCWISLKLKCVLIIF  
FGVYDAGDLLFLPIPERFIHQMDRINEATSVLSKRWRYLFAFRPNCLDDQEE  
VGGDSFIDFVDRVLVTGTFPIRIRISIKRMSIDTGHVTRMVDVLEHGVSYLDIDI  
TSEDDIGFVPLEFTCKTLVELKLARGFYAPVDVYSLPSLKTLSLWFCNRDCCP  
LGRLLSACPVELEHTIGGCWQHIEFCRTVSSVSKLKLITSSFYHWGIIIDTPSL  
AYLESDLVPRHPFNLESLEAKLALDCLSDSNPTNLIKGLNVEYVLELSSGXTSQ  
IILYDREAPVLKSLFRLSKNSRHYIWNLLPILLESKPLETLIVKGLPSADRYE  
RHYGLSCPVKVLEIYEGGKYEELEQMEHFLKPLCELVLKVRASAINDKERSRITKD  
LUMVPRSSNCNKLKFC"  
complement(13016..13080)  
/number=1  
complement(13081..13227)  
/gene="AT4g22280"  
/number=2  
complement(13228..13378)  
/number=2  
complement(13379..14157)  
/gene="AT4g22280"  
/number=3  
complement(14158..14208)  
/number=3  
complement(14209..14314)  
/gene="AT4g22280"  
/number=4  
complement(14315..14994)  
/number=4  
complement(14995..15111)  
/gene="AT4g22280"  
/number=5

Query Match 81.7%; Score 18.8; DB 8; Length 199577;  
Best Local Similarity 90.9%; Pred. No. 1e+02; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 2

QY 2 aaatccatgtaaacgacgagg 23  
|||||  
Db 159343 AAATCCATGTAAAGCAGTAGG 159322

RESULT 8  
AC072048/c  
LOCUS  
DEFINITION Mus musculus chromosome 6 clone RP23-189E15 strain C57BL6/J,  
WORKING DRAFT SEQUENCE, 24 unordered pieces.  
ACCESSION AC072048  
VERSION AC072048.1 GI:8313202  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 208684)  
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,  
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,  
Huang,M.C., Idol,J., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B.,  
Mastrian,S., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,  
Stanton,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,  
Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,  
Wetherby,K.D. and Green,E.D.  
NISC Mouse Sequencing Initiative  
Unpublished  
2 (bases 1 to 208684)  
Green,E.D.  
Direct Submission  
Submitted (07-JUN-2000) NIH Intramural Sequencing Center, 8717  
Grovemont Circle, Gaithersburg, MD 20877, USA  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc\_mouse@nhgri.nih.gov  
----- Project Information  
Center project name: wp  
Center clone name: 189E15  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 195465 bases at least Q40  
Consensus quality: 197294 bases at least Q30  
Consensus quality: 198446 bases at least Q20  
Insert size: 209000; agarose-fp  
Insert size: 197000; pulse-field-gel  
Insert size: 206384; sum-of-contigs  
Quality coverage: 6.43x in Q20 bases; agarose-fp  
Quality coverage: 6.82x in Q20 bases; pulse-field-gel  
Quality coverage: 6.51x in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 24 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2209: contig of 2209 bp in length  
\* 2210 2309: gap of unknown length  
\* 2310 4736: contig of 2427 bp in length  
\* 4737 4836: gap of unknown length  
\* 4837 7023: contig of 2187 bp in length  
\* 7024 7123: gap of unknown length  
\* 7124 9505: contig of 2382 bp in length  
\* 9506 9605: gap of unknown length  
\* 9606 14706: contig of 5101 bp in length  
\* 14707 14806: gap of unknown length  
\* 14807 18659: contig of 3853 bp in length  
\* 18660 18759: gap of unknown length  
\* 18760 25424: contig of 6665 bp in length  
\* 25425 25524: gap of unknown length  
\* 25525 33299: contig of 7775 bp in length  
\* 33300 33399: gap of unknown length  
\* 33400 39449: contig of 6050 bp in length  
\* 39450 39549: gap of unknown length  
\* 39550 48515: contig of 8966 bp in length  
\* 48516 48615: gap of unknown length  
\* 48616 55023: contig of 6408 bp in length  
\* 55024 55123: gap of unknown length  
\* 55124 64442: contig of 9319 bp in length  
\* 64443 64542: gap of unknown length  
\* 64543 73027: contig of 8485 bp in length  
\* 73028 73127: gap of unknown length  
\* 73128 82147: contig of 9020 bp in length  
\* 82148 82247: gap of unknown length

```

* 82248 93854: contig of 11607 bp in length
* 93855 93954: gap of unknown length
* 103236 93955: contig of 9282 bp in length
* 103237 103336: gap of unknown length
* 103337 117502: contig of 14166 bp in length
* 117503 117602: gap of unknown length
* 117603 130162: contig of 12560 bp in length
* 130163 130262: gap of unknown length
* 130263 140708: contig of 10446 bp in length
* 140709 140808: gap of unknown length
* 140809 151652: contig of 10844 bp in length
* 151653 151752: gap of unknown length
* 151753 163824: contig of 12071 bp in length
* 163824 163923: gap of unknown length
* 163924 178730: contig of 14807 bp in length
* 178731 178830: gap of unknown length
* 178831 193909: contig of 15079 bp in length
* 193910 194009: gap of unknown length
* 194010 208684: contig of 14675 bp in length.
FEATURES
    source
        1. .208684
            /organism="Mus musculus"
            /strain="C57BL6/J"
            /db_xref="taxon:10090"
            /chromosome="6"
            /clone_lib="RPCI mouse BAC library 23"
            1. .2209
                /note="assembly_fragment"
                2310. 4736
                    /note="assembly_fragment"
                    4837. .7023
                        /note="assembly_fragment"
                        7124. .9505
                            /note="assembly_fragment"
                            9606. 14706
                                /note="assembly_fragment"
                                14807. .18659
                                    /note="assembly_fragment"
                                    18760. .25424
                                        /note="assembly_fragment"
                                        25525. .33299
                                            /note="assembly_fragment"
                                            33400. .39449
                                                /note="assembly_fragment"
                                                39550. .48515
                                                    /note="assembly_fragment"
                                                    48616. .55023
                                                        /note="assembly_fragment"
                                                        55124. 64442
                                                            /note="assembly_fragment"
                                                            64543. .73027
                                                                /note="assembly_fragment"
                                                                73128. .82147
                                                                    /note="assembly_fragment"
                                                                    82248. .93854
                                                                        /note="assembly_fragment"
                                                                        93955. .103236
                                                                            /note="assembly_fragment"
                                                                            103337. .117502
                                                                                /note="assembly_fragment"
                                                                                117603. .130162
                                                                                    /note="assembly_fragment"
                                                                                    130263. .140708
                                                                                        /note="assembly_fragment"
                                                                                        140809. .151652
                                                                                            /note="assembly_fragment"
                                                                                            151753. .163823
                                                                                                /note="assembly_fragment"
                                                                                                163924. 178730
                                                                                                    /note="assembly_fragment"
                                                                                                    clone_end:r7
                                                                                                        vector_side:right"

```

## misc\_feature

```

178831. .193909
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
194010. .208684
/note="assembly_fragment"

```

## misc\_feature

```

BASE COUNT 54390 a 48824 c 48605 g 54542 t 2323 others
ORIGIN

```

## Query Match

```

Best Local Similarity 81.7%; Score 18.8; DB 2; Length 208684;
Matches 20; Conservative 0; Pred.No.1e+02; 2; Indels 0; Gaps 0;

```

```

QY 2 aaatcatgttaagcagcagg 23
||||| ||||| ||||| |||||

```

```

Db 135605 AAATCCTGTAACACGACGAGG 135584

```

## RESULT 9

```

AC046148
LOCUS

```

```

DEFINITION
Mus musculus chromosome 14 clone RP23-265D21, *** SEQUENCING IN
PROGRESS ***, 45 unordered pieces.
ACCESSION
AC046148
VERSION
AC046148.5 GI:11094639
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 168524)
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, X., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Oguh, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,
Worley, K. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 168524)
Worley, K.C.
Direct Submission
Submitted (13-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 4, 2000 this sequence version replaced gi:9929612.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAEB
Center clone name: Rp23-265D21
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 132230 bases at least Q40
Consensus quality: 149333 bases at least Q30
Consensus quality: 156096 bases at least Q20
Estimated insert size: 153798; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-1p estimation

```

Quality coverage: 2.5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 10853: contig of 10853 bp in length  
\* 10854 10953: gap of unknown length  
\* 10954 20346: contig of 9393 bp in length  
\* 20347 20446: gap of unknown length  
\* 20447 30391: contig of 9945 bp in length  
\* 30392 30491: gap of unknown length  
\* 30492 39349: contig of 8858 bp in length  
\* 39350 39449: gap of unknown length  
\* 39450 47183: contig of 7734 bp in length  
\* 47184 53762: contig of 6479 bp in length  
\* 53763 53862: gap of unknown length  
\* 53863 58689: contig of 4807 bp in length  
\* 58690 58769: gap of unknown length  
\* 58770 65058: contig of 6289 bp in length  
\* 65059 69830: contig of 4672 bp in length  
\* 69831 73910: gap of unknown length  
\* 73911 74010: gap of unknown length  
\* 74011 78684: contig of 4674 bp in length  
\* 78685 78784: gap of unknown length  
\* 78785 83343: contig of 4559 bp in length  
\* 83344 83444: gap of unknown length  
\* 83444 87849: contig of 4406 bp in length  
\* 87850 87949: gap of unknown length  
\* 87950 92076: contig of 4127 bp in length  
\* 92077 92176: gap of unknown length  
\* 92177 97227: contig of 5051 bp in length  
\* 97228 97327: gap of unknown length  
\* 97328 101356: contig of 4029 bp in length  
\* 101357 101456: gap of unknown length  
\* 101457 104693: contig of 3237 bp in length  
\* 104694 104793: gap of unknown length  
\* 104794 108132: contig of 3339 bp in length  
\* 108133 108232: gap of unknown length  
\* 108233 111747: contig of 3515 bp in length  
\* 111748 111847: gap of unknown length  
\* 111848 115180: contig of 3333 bp in length  
\* 115181 115280: gap of unknown length  
\* 115281 118548: contig of 3268 bp in length  
\* 118549 118648: gap of unknown length  
\* 118649 121539: contig of 2891 bp in length  
\* 121540 121639: gap of unknown length  
\* 121640 124526: contig of 2887 bp in length  
\* 124527 124626: gap of unknown length  
\* 124627 127512: contig of 2886 bp in length  
\* 127513 127612: gap of unknown length  
\* 127613 131027: contig of 3415 bp in length  
\* 131028 131127: gap of unknown length  
\* 131128 133604: contig of 2477 bp in length  
\* 133605 133704: gap of unknown length  
\* 133705 135894: contig of 2190 bp in length  
\* 135895 135994: gap of unknown length  
\* 135995 137407: contig of 1413 bp in length  
\* 137408 137507: gap of unknown length  
\* 137508 139522: contig of 2015 bp in length  
\* 139523 139622: gap of unknown length  
\* 139623 141182: contig of 1560 bp in length  
\* 141183 141283: gap of unknown length  
\* 141283 143104: contig of 1822 bp in length

\* 143105 143204: gap of unknown length  
\* 143205 145597: contig of 2393 bp in length  
\* 145598 145697: gap of unknown length  
\* 145698 147840: contig of 2143 bp in length  
\* 147841 147940: gap of unknown length  
\* 147941 150388: contig of 2448 bp in length  
\* 150389 150488: gap of unknown length  
\* 150489 151853: contig of 1365 bp in length  
\* 151854 151953: gap of unknown length  
\* 151954 154894: contig of 2941 bp in length  
\* 154895 154994: gap of unknown length  
\* 154995 156600: contig of 1606 bp in length  
\* 156601 156700: gap of unknown length  
\* 156701 158135: contig of 1435 bp in length  
\* 158136 158235: gap of unknown length  
\* 158236 159267: contig of 1032 bp in length  
\* 159268 159367: gap of unknown length  
\* 159368 161428: contig of 2061 bp in length  
\* 161429 161528: gap of unknown length  
\* 161529 162993: contig of 1465 bp in length  
\* 162994 163093: gap of unknown length  
\* 163094 164387: contig of 1293 bp in length  
\* 164387 164887: gap of unknown length  
\* 164887 165850: contig of 1364 bp in length  
\* 165851 165950: gap of unknown length  
\* 165951 167364: contig of 1414 bp in length  
\* 167365 167465: gap of unknown length  
\* 167465 168524: contig of 1060 bp in length.

FEATURES  
source

1..168524  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="14"  
/clone="RP23-265D21"  
BASE COUNT 45461 a 36455 c 36516 g 45685 t 4407 others  
ORIGIN

Query Match 80.0%; Score 18.4; DB 2; Length 168524;  
Best Local Similarity 95.0%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aaatccatgtaagcagcag 21  
|||||  
Db 112558 AATCCATGTAACACGACG 112577

RESULT 10  
AC040920/c

LOCUS AC040920 172650 bp DNA HTG 11-APR-2000  
DEFINITION Homo sapiens chromosome 5 clone RP11-509M23 map 5, WORKING DRAFT  
SEQUENCE, 19 unordered pieces.

ACCESSION AC040920

VERSION AC040920.1 GI:7534094

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 172650)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 172650)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,

Boguslavsky,L., Bouckhalter,B., Brown,A., Burckett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrum, J., Meneus, L., Milova, I., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Vleel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

# TITLE JOURNAL

## COMMENT

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L8553

Center clone name: 509\_M\_23

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 163069 bases at least Q40

Consensus quality: 167976 bases at least Q30

Consensus quality: 169595 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 170850; sum-of-ontigs

Quality coverage: 4.7 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 bases; sum-of-ontigs

----- NOTE: This is a 'working draft' sequence. It currently  
 consists of 19 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 422: contig of 422 bp in length  
 423 522: gap of 100 bp  
 523 2269: contig of 1747 bp in length  
 2270 2369: gap of 100 bp  
 2370 4946: contig of 2577 bp in length  
 4947 5046: gap of 100 bp  
 5047 7262: contig of 2116 bp in length  
 7163 9787: contig of 2525 bp in length  
 9788 9887: gap of 100 bp  
 9888 13694: contig of 3807 bp in length  
 13695 13794: gap of 100 bp  
 13795 18482: contig of 4688 bp in length  
 18483 18582: gap of 100 bp  
 18583 24380: contig of 5798 bp in length  
 24381 24480: gap of 100 bp  
 24481 30597: contig of 6117 bp in length  
 30598 30697: gap of 100 bp  
 30698 36241: contig of 5544 bp in length  
 36242 36341: gap of 100 bp  
 36342 43886: contig of 7545 bp in length  
 43887 43986: gap of 100 bp  
 43987 53642: contig of 9656 bp in length  
 53643 53742: gap of 100 bp  
 53743 65843: contig of 12101 bp in length  
 65844 65943: gap of 100 bp  
 65944 82865: contig of 16922 bp in length

\* 82866 82965: gap of 100 bp  
 82966 97230: contig of 14265 bp in length  
 97231 97330: gap of 100 bp  
 97331 114682: contig of 17352 bp in length  
 114683 114782: gap of 100 bp  
 114783 132289: contig of 17507 bp in length  
 132290 132389: gap of 100 bp  
 132390 151664: contig of 19275 bp in length  
 151665 151764: gap of 100 bp  
 151765 172650: contig of 20886 bp in length.

## FEATURES

### source

Location/Qualifiers  
 1. .172650  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /map="5"  
 /clone="RP11-509M23"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. .422  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:left  
 523. 2269  
 /note="assembly\_fragment"  
 2370. .4946  
 /note="assembly\_fragment"  
 5047. .7162  
 /note="assembly\_fragment"  
 7263. .9787  
 /note="assembly\_fragment"  
 9888. .13694  
 /note="assembly\_fragment"  
 13795. .18482  
 /note="assembly\_fragment"  
 18583. .24380  
 /note="assembly\_fragment"  
 24481. .30597  
 /note="assembly\_fragment"  
 30698. .36241  
 /note="assembly\_fragment"  
 36342. .43886  
 /note="assembly\_fragment"  
 43987. .53642  
 /note="assembly\_fragment"  
 53743. .65843  
 /note="assembly\_fragment"  
 65944. .82865  
 /note="assembly\_fragment"  
 82966. .97230  
 /note="assembly\_fragment"  
 97331. .114682  
 /note="assembly\_fragment"  
 114783. .132289  
 /note="assembly\_fragment"  
 clone\_end:Sp6  
 vector\_side:left  
 132390. .151664  
 /note="assembly\_fragment"  
 151765. .172650  
 /note="assembly\_fragment"  
 1804 others

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

### misc\_feature

misc\_feature

Query Match 80.0%; Score 18.4; DB 2; Length 172650;  
 Best Local Similarity 95.0%; Pred. No. 1.6e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 2 aaatccatgtaaacgacag 21  
 ||| |||||

Db 142729 AAAGCCATGTAAAGCAGCAG 142710

```
RESULT 11
AF118429
LOCUS      AF118429      535 bp      DNA      VRT      10-JAN-2001
DEFINITION Falco peregrinus microsatellite NVH Fp86-2 sequence.
ACCESSION  AF118429
VERSION     AF118429.1  GI:6840932
KEYWORDS
SOURCE      Falco peregrinus.
ORGANISM    Falco peregrinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Falconiformes; Falconidae; Falco.
REFERENCE   1 (bases 1 to 535)
AUTHORS    Nesje,M., Roed,K.H., Lifjeld,J.T., Lindberg,P. and Steen,O.F.
TITLE      Genetic relationships in the peregrine falcon (Falco peregrinus)
            analysed by microsatellite DNA markers
JOURNAL    Mol. Ecol. 9 (1), 53-60 (2000)
MEDLINE    20117586
PUBMED     10652075
REFERENCE   2 (bases 1 to 535)
AUTHORS    Nesje,M. and Roed,K.H.
TITLE      Direct Submission
JOURNAL    Submitted (07-JAN-1999) MGA, Genetics, Norwegian College of
            Veterinary Medicine, P.O. Box 8146 Dep., Oslo 0033, Norway
FEATURES
source      Location/Qualifiers
            1..535
            /organism="Falco peregrinus"
            /db_xref="taxon:8954"
repeat_region 1..535
            /note="microsatellite NVH Fp86-2"
            /rpt_type=tandem
BASE COUNT  168 a 110 c 117 g 140 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 5; Length 535;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgagg 23
   | ||||| ||||| ||||| |||||
Db 385 GGAATCCATGACAGCAGGG 407

RESULT 12
AC005083/c
LOCUS      AC005083      146285 bp      DNA      PRI      21-DEC-1999
DEFINITION Homo sapiens BAC clone CTA-281G5 from 7p15-p21, complete sequence.
ACCESSION  AC005083
VERSION     AC005083.1  GI:4150930
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 146285)
AUTHORS    Madsen,C. and Blair,T.
TITLE      The sequence of Homo sapiens BAC clone CTA-281G5
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 146285)
AUTHORS    Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3 (bases 1 to 146285)
AUTHORS    Waterston,R.
TITLE      Direct Submission
JOURNAL    Submitted (12-JAN-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4 (bases 1 to 146285)
AUTHORS    Waterston,R.
TITLE      Direct Submission
JOURNAL    Submitted (21-DEC-1999) Department of Genetics, Washington

COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_RG281G05
-----

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 12, 1999 this sequence version replaced gi:3212908.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_RG281G05
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7 , send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
Clone CTA-281G5 is from a release of the human BAC library
CITB-HS-A. The library contains cloned DNA from human sperm. See:
Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
Kim et al., Genomics 34:213-8 (1996). The clone is available from
Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of CTA-281G5;
actual end is at 146285 of CTA-281G5

BAC CTA-281G5 contains an E. coli transposon from 10982 to 12345
that is not represented in the submitted sequence.

FEATURES
source      1..146285
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="7"
            /map="7p15-p21"
            /clone="CTA-281G5"
            /clone_lib="CITB-HS-A"
            322..358
            /rpt_family="(TAAAA)n"
            808..952
            /rpt_family="MER1_type"
            1172..1194
            /rpt_family="AT_rich"
            1264..1720
            /rpt_family="L1"
            1848..1881
            /rpt_family="AT_rich"
            1882..2171
            /rpt_family="Alu"
            3782..3804
            /rpt_family="AT_rich"
            4794..4826
            /rpt_family="AT_rich"
            5764..6034
```

```
repeat_region /rpt_family="purine-rich"
6891. 7126 /rpt_family="L1"
repeat_region /rpt_family="L1"
7133. 7263 /rpt_family="L1"
repeat_region /rpt_family="L1"
7455. 7481 /rpt_family="L1"
repeat_region /rpt_family="AT_rich"
7504. 7554 /rpt_family="AT_rich"
repeat_region /rpt_family="AT_rich"
8320. 8348 /rpt_family="AT_rich"
repeat_region /rpt_family="CA)n"
8466. 8533 /rpt_family="CA)n"
repeat_region /rpt_family="AT_rich"
9611. 9733 /rpt_family="AT_rich"
repeat_region /rpt_family="MIR"
9946. 10172 /rpt_family="MIR"
repeat_region /rpt_family="MERL_type"
10437. 10477 /rpt_family="MERL_type"
repeat_region /rpt_family="CA)n"
10581. 10681 /rpt_family="CA)n"
repeat_region /rpt_family="(TAAA)n"
10759. 11049 /rpt_family="(TAAA)n"
repeat_region /rpt_family="AT_rich"
11050. 11235 /rpt_family="AT_rich"
repeat_region /rpt_family="(CAATA)n"
11408. 11526 /rpt_family="(CAATA)n"
repeat_region /rpt_family="(CATAT)n"
11768. 11872 /rpt_family="(CATAT)n"
repeat_region /rpt_family="(CATTA)n"
12002. 12110 /rpt_family="(CATTA)n"
repeat_region /rpt_family="(CATAT)n"
12057. 12168 /rpt_family="(CATAT)n"
repeat_region /rpt_family="(TA)n"
12381. 12490 /rpt_family="(TA)n"
repeat_region /rpt_family="(CAATA)n"
12562. 12680 /rpt_family="(CAATA)n"
repeat_region /rpt_family="(CTATA)n"
13131. 13367 /rpt_family="(CTATA)n"
repeat_region /rpt_family="L2"
13378. 13540 /rpt_family="L2"
repeat_region /rpt_family="MERL_type"
13557. 13744 /rpt_family="MERL_type"
repeat_region /rpt_family="L2"
13745. 14592 /rpt_family="L2"
repeat_region /rpt_family="L1"
14593. 14616 /rpt_family="L1"
repeat_region /rpt_family="L2"
16078. 16348 /rpt_family="L2"
repeat_region /rpt_family="L1"
18492. 18692 /rpt_family="L1"
repeat_region /rpt_family="MERL_type"
19334. 19422 /rpt_family="MERL_type"
repeat_region /rpt_family="LTR41"
19880. 20137 /rpt_family="LTR41"
repeat_region /rpt_family="LTR41"
20504. 20541 /rpt_family="LTR41"
repeat_region /rpt_family="MERL_type"
21092. 21607 /rpt_family="MERL_type"
repeat_region /rpt_family="MaLR"
22221. 22499 /rpt_family="MaLR"
repeat_region /rpt_family="Alu"
22500. 22523 /rpt_family="Alu"
repeat_region /rpt_family="CAAAA)n"
22535. 22857 /rpt_family="CAAAA)n"
repeat_region /rpt_family="MER4-group?"
23203. 23380 /rpt_family="MER4-group?"
repeat_region /rpt_family="L1"
23381. 23643 /rpt_family="L1"
repeat_region /rpt_family="Alu"
23644. 24770 /rpt_family="Alu"
repeat_region /rpt_family="L1"
25826. 25846 /rpt_family="L1"
repeat_region /rpt_family="AT_rich"
```

```
repeat_region 26130. 26254 /rpt_family="MIR"
repeat_region 26529. 26563 /rpt_family="MIR"
repeat_region 26792. 27116 /rpt_family="AT_rich"
repeat_region 30744. 30982 /rpt_family="Alu"
repeat_region 31032. 31739 /rpt_family="MIR"
repeat_region 32332. 32448 /rpt_family="L1"
repeat_region 32536. 32730 /rpt_family="MIR"
repeat_region 33282. 33333 /rpt_family="MIR"
repeat_region 35261. 35288 /rpt_family="AT_rich"
repeat_region 35571. 35598 /rpt_family="(CATTT)n"
repeat_region 35924. 36007 /rpt_family="(CA)n"
repeat_region 36476. 37873 /rpt_family="L2"
repeat_region 38324. 38899 /rpt_family="L1"
repeat_region /rpt_family="Retroviral"
```

Query Match 79.1%; Score 18.2; DB 9; Length 146285;  
Best Local Similarity 87.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaatccatgttaagcagcaggg 23

Db 20614 GACATCCATGTAAAGGACGAGG 20592

RESULT 13  
AC023167/c

LOCUS AC023167 183634 bp DNA HTG 04-NOV-2000  
DEFINITION Mus musculus clone RP23-361K18, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 61  
unordered pieces.

ACCESSION AC023167  
VERSION AC023167.8 GI:11079356

KEYWORDS HTG; HTGS, PHASE1.

SOURCE house mouse

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 183634)  
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,  
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,  
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,  
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,  
Fernandez, C., Ferraguto, P., Forcum-Tansey, J., Gill, R.,  
Gorell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,  
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,  
Kovar, C., Liu, J., Liu, W., Louisedge, H., Lozano, R. J., Martin, R.,  
Massey, E., McLeod, M. P., Mei, G., Moore, S., Morgan, M., Morris, S.,  
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,  
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,  
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,  
Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,  
Worley, K. and Gibbs, R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 183634)

AUTHORS Worley, K.C.

Direct Submission

Submitted (09-FEB-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 3, 2000 this sequence version replaced gi:8248589.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: RP23-361K18
Center clone name: RP23-361K18
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-primer Bodipy: 70% of reads
Chemistry: Dye-terminator Big Dye: 30% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 137454 bases at least Q40
Consensus quality: 161084 bases at least Q30
Consensus quality: 170194 bases at least Q20
Estimated insert size: 164611; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 10583: contig of 10583 bp in length
* 10584 10683: gap of unknown length
* 10684 20427: contig of 9744 bp in length
* 20428 20527: gap of unknown length
* 20528 29063: contig of 8536 bp in length
* 29064 29163: gap of unknown length
* 29164 36211: contig of 7048 bp in length
* 36212 36311: gap of unknown length
* 36312 42254: contig of 5943 bp in length
* 42255 42354: gap of unknown length
* 42355 47592: contig of 5338 bp in length
* 47592 47792: gap of unknown length
* 47793 51859: contig of 4067 bp in length
* 51860 51959: gap of unknown length
* 51960 55000: contig of 3041 bp in length
* 55001 55101: gap of unknown length
* 55101 59766: contig of 4666 bp in length
* 59767 59866: gap of unknown length
* 59867 62657: contig of 2791 bp in length
* 62658 62757: gap of unknown length
* 62758 67170: contig of 4413 bp in length
* 67171 67270: gap of unknown length
* 67271 70199: contig of 2929 bp in length
* 70200 70299: gap of unknown length
* 70300 74125: contig of 3826 bp in length
* 74126 74225: gap of unknown length
* 74226 74307: contig of 3082 bp in length
* 74308 77407: gap of unknown length
* 77408 79603: contig of 2196 bp in length
* 79604 79703: gap of unknown length
* 79704 82685: contig of 2982 bp in length
* 82686 82785: gap of unknown length
* 82786 86384: contig of 3599 bp in length
* 86385 86484: gap of unknown length
* 86485 90067: contig of 3583 bp in length
* 90068 90167: gap of unknown length
* 90168 93018: contig of 2851 bp in length
* 93019 93118: gap of unknown length
* 93119 96993: contig of 3875 bp in length
* 96994 97093: gap of unknown length
* 97094 100012: contig of 2919 bp in length
* 100013 100112: gap of unknown length
* 100113 103857: contig of 3745 bp in length
*
* 103858 103957: gap of unknown length
* 103958 107522: contig of 3565 bp in length
* 107523 110046: gap of unknown length
* 110047 110146: gap of unknown length
* 110147 113098: contig of 2952 bp in length
* 113099 113199: gap of unknown length
* 113199 116570: contig of 3372 bp in length
* 116571 116670: gap of unknown length
* 116671 119013: contig of 2343 bp in length
* 119014 119113: gap of unknown length
* 119114 122053: contig of 2940 bp in length
* 122054 122153: gap of unknown length
* 122154 124959: contig of 2806 bp in length
* 124960 125059: gap of unknown length
* 125060 127168: contig of 2109 bp in length
* 127169 127288: gap of unknown length
* 127289 129604: contig of 2336 bp in length
* 129605 129704: gap of unknown length
* 129705 133155: contig of 3451 bp in length
* 133156 133255: gap of unknown length
* 133256 136608: contig of 3353 bp in length
* 136609 136708: gap of unknown length
* 136709 138127: contig of 1419 bp in length
* 138128 138227: gap of unknown length
* 138228 141003: contig of 2776 bp in length
* 141004 141103: gap of unknown length
* 141104 143694: contig of 2591 bp in length
* 143695 143794: gap of unknown length
* 143795 144911: contig of 1117 bp in length
* 144912 145011: gap of unknown length
* 145012 146360: contig of 1349 bp in length
* 146361 146460: gap of unknown length
* 146461 148990: contig of 2530 bp in length
* 148991 149090: gap of unknown length
* 149091 150392: contig of 1302 bp in length
* 150393 150492: gap of unknown length
* 150493 152202: contig of 1710 bp in length
* 152203 152303: gap of unknown length
* 152304 154394: contig of 2092 bp in length
* 154395 154494: gap of unknown length
* 154495 156415: contig of 1921 bp in length
* 156416 156515: gap of unknown length
* 156516 157973: contig of 1458 bp in length
* 157974 158073: gap of unknown length
* 158074 159544: contig of 1471 bp in length
* 159545 159644: gap of unknown length
* 159645 161111: contig of 1467 bp in length
* 161112 162823: contig of 1612 bp in length
* 162824 162923: gap of unknown length
* 162924 164564: contig of 1641 bp in length
* 164565 164664: gap of unknown length
* 164665 166095: contig of 1431 bp in length
* 166096 166195: gap of unknown length
* 166196 167971: contig of 1776 bp in length
* 167972 168071: gap of unknown length
* 168072 169306: contig of 1235 bp in length
* 169307 169406: gap of unknown length
* 169407 170694: contig of 1288 bp in length
* 170695 170794: gap of unknown length
* 170795 172353: contig of 1559 bp in length
* 172354 172453: gap of unknown length
* 172454 173463: contig of 1010 bp in length
* 173464 173563: gap of unknown length
* 173564 174902: contig of 1339 bp in length
* 174903 175002: gap of unknown length
* 175003 176926: contig of 1924 bp in length
* 176927 177026: gap of unknown length
* 177027 178509: contig of 1483 bp in length
* 178510 178609: gap of unknown length
* 178610 179834: contig of 1225 bp in length
* 179835 179934: gap of unknown length

```





misc\_feature /map="11"  
/clone="RP11-692M12"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 3785  
clone\_end:SP6  
vector\_side:left  
3886. .5107  
/note="assembly\_fragment"  
misc\_feature 5208. .6375  
/note="assembly\_fragment"  
misc\_feature 6476. .7525  
/note="assembly\_fragment"  
misc\_feature 7626. .8701  
/note="assembly\_fragment"  
misc\_feature 8802. .10296  
/note="assembly\_fragment"  
misc\_feature 10397. .11470  
/note="assembly\_fragment"  
misc\_feature 11571. .12684  
/note="assembly\_fragment"  
misc\_feature 12785. .14999  
/note="assembly\_fragment"  
misc\_feature 15100. .18672  
/note="assembly\_fragment"  
misc\_feature 18773. .22303  
/note="assembly\_fragment"  
misc\_feature 22404. .28525  
/note="assembly\_fragment"  
misc\_feature 28626. .36093  
/note="assembly\_fragment"  
misc\_feature 36194. .45198  
/note="assembly\_fragment"  
misc\_feature 45299. .56916  
/note="assembly\_fragment"  
misc\_feature 57017. .65985  
/note="assembly\_fragment"  
misc\_feature 66086. .78887  
/note="assembly\_fragment"  
misc\_feature 78988. .94277  
/note="assembly\_fragment"  
misc\_feature 94378. .10994  
/note="assembly\_fragment"  
misc\_feature 110095. .121149  
/note="assembly\_fragment"  
misc\_feature 121250. .138125  
/note="assembly\_fragment"  
misc\_feature 138226. .155392  
/note="assembly\_fragment"  
misc\_feature 155493. .179217  
/note="assembly\_fragment"  
misc\_feature 179318. .202705  
/note="assembly\_fragment"  
misc\_feature 202806. .228809  
/note="assembly\_fragment"  
BASE COUNT 62480 a 50546 c 48170 g 65210 t 2403 others  
ORIGIN

Query Match 79.1%; Score 18.2; DB 2; Length 228809;  
Best local Similarity 87.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgagg 23  
||||||| | ||||| |||||  
Db 97503 GAATCCAGCAAGCATCAGGG 97481

RESULT 15  
AC007294 AC007294 23486 bp DNA PLN 05-APR-2000  
LOCUS Arabidopsis thaliana chromosome II section 75 of 255 of the  
DEFINITION complete sequence. Sequence from clones T26C18, T10F5.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC007294 AB002093  
AC007294.7 GI:6598756  
HTG.  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
1 (bases 1 to 23486)  
Lin.X., Kaul.S., Rounsley.S.D., Shea.T.P., Benito.M.-I., Town.C.D.,  
Fujii.C.Y., Mason.T.M., Bowman.C.L., Barnstead.M.E., Ronning.C.M.,  
Feldblum.T.V., Lee.J.C., Cronin.L.A., Shen.M., VanAken.S.E., Umayam.L.,  
Koo.H., Moffat.K.S., Cronin.L.A., Carrera.A.J., Creasy.T.H.,  
Tallon.L.J., Gill.J.E., Adams.M.D., Somerville.C.R., Copenhaver.G.P., Preuss.D.,  
Goodman.H.M., Somerville.C.R., Copenhaver.G.P., Preuss.D.,  
Nierman.W.C., White.O., Eisen.J.A., Salzberg.S.L., Fraser.C.M. and  
Venter.J.C.  
Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana  
Nature 402 (6763), 761-768 (1999)  
20083487  
10617197  
2 (bases 1 to 23486)  
Lin.X.  
Direct Submission  
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
On Dec 17, 1999 this sequence version replaced gi:4733957.  
The sequence and annotation of chromosome 2 were merged from those  
of the individual clones on this chromosome after removing  
overlaps. For detailed information, please see the TIGR web site  
(http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene  
prediction programs including GRAIL  
(ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,  
University of Washington), Genscan (Chris Burge,  
http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene  
(http://www.cbs.dtu.dk/services/NetGene2/), searches of the  
complete sequence against a peptide database and plant EST  
databases at TIGR, and manual curations based on those analyses.  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as 'unknown' proteins.  
Genes without protein or EST similarity, that are predicted by two  
or more gene prediction programs over most of their length are  
annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
annotated as 'tRNA-scan-SE' (Sean Eddy,  
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were  
identified by repeatmasker (Arian Smit,  
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are  
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones  
F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone  
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards  
and Satoshi Tabata for helpful assistance. In addition, we would  
like to thank the TIGR Bioinformatics Department, especially Lixin  
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy  
Peterson, Michael Holmes, and Delwood Richardson for software and  
database support.

This work was supported by the National Science Foundation,  
Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES  
Source  
Location/Qualifiers  
1. 23486  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="II"

```

misc_feature
/1..23285
/note="Sequence from clone T26C18"
complement(65..99)
/rpt_family="(CAAAA)n"
complement(261..1274)
/gene="At2g13420"
/note="At2g13420"
/pseudo
complement(2460..2567)
/rpt_family="(CAT)n"
2738..2818
/rpt_family="(GA)n"
complement(2857..2950)
/rpt_family="POLY_A"
join(<3764..3786,4264..4517,4869..5134)
/gene="At2g13430"
<3764..5134
/gene="At2g13430"
/note="T26C18.2"
join(3764..3786,4264..4517,4869..5134)
/gene="At2g13430"
/note="unknown protein"
/codon_start=1
/protein_id="AAD28642.1"
/db_xref="GI:4733958"
/translation="MKIDNLYIFHLNQMPAAACAGKELKSSDLERFKAWSG
FYDWFORISANNRNRLTTAKRVLAGPKILQTQIOVKCSPINANSRQRSGRSNK
ASHERNRTNRKNSKHEEANGKDKPTIAALKAIIDGRELANTTKVRSCQTISPTT
RDMETGKTFGQENSSTFSL"
complement(join(<6982..7176,7257..7349,7438..7661,
7805..7971,8467..8609,8779..8997,9332..9412,10180..10359,
10550..10880,11089..11202,11434..11525,11601..11737,
11813..>12008))
/gene="At2g13440"
complement(<6982..>12008)
/gene="At2g13440"
/note="T26C18.3; similar to glucose inhibited division
protein A from prokaryotes; contains a glucose inhibited
division protein A family signature (PF:PF01134)"
7805..7971,8467..8609,8779..8997,9332..9412,10180..10359,
10550..10880,11089..11202,11434..11525,11601..11737,
11813..12008)
/gene="At2g13440"
/codon_start=1
/product="GIDA-like protein"
/protein_id="AAD28643.1"
/db_xref="GI:4733959"
/translation="MRAAAATATVSLRHRSPPTIVPSLLFSSSSIPFHPRLCVF
LRPRQLNRLPAASFSSSSGATSDTVDVIVVAGHAGCAALASARLGASTLLLT
LNLDRIAWQPCNPAGVGPAGKASOLVHEVDALGGDVGKADRCYLQKRILNVSGPAVRS
LRAQDKREVATEMKKIVDSTENLCIREAWVDIIVGKNDVEGVATFGMNFYAPSV
ILTGTFMGKIWVGKSKMPAGPAGESASQGLTENIQKLGFDRLKLTGPARDVRT
IDFNLGAOGBDEVSFDFPDFFHIERQMCCTLTTRTKITHQIRNLHETPTYGG
WVAKGPRCPSTEDKIVRKOKESHQIFLEPEGRDVPPIYVQGFSTGLPENLQPLL
RSLPGENSMRLPAYBYDYLPAPHCSSRLMTKKIGLFPSGOINGTTGYEAAAO
GIISGNAARHADGKKHVYLERESSYIGTLDLVTKDLREPYRMLTSRSEHRLIRE
DNADSLTPIGRELGLIDDRRWKLQKQOARISEBKRLKTVKISVAVGDLAAEVSSV
SSQPVKESATLLESLLKPHIKLEKHGFGNETLSRMEKDCVEIDIKYGFIVRQON
QLQOMVHOORHRLPDDLVDYSMTLSHGREKLSKVRPETIGQASRVGGVSPADITAL
LITLESNRRTQDVRGKILEHALAESNPQWVEDREHVYNE"
11854..11935
/rpt_family="(GGA)n"
complement(12073..12132)
/rpt_family="POLY_A"
complement(<12451..>13635)
/gene="At2g13450"
complement(<12451..>13635)
/gene="At2g13450"
/note="T26C18.4; similar to GB:AAA75253; some members of
this protein family have a weak CCHC zinc fingers that is
mostly from retroviral gag proteins (nucleocapsid)"
complement(12451..13635)
CDS

```

```

/gene="At2g13450"
/codon_start=1
/product="putative Tall-like non-LTR retroelement protein"
/protein_id="AAD28644.1"
/db_xref="GI:4733960"
/translation="MADELWDELQHLGLREDPALFIPHEAYATIVESRNLISLIARPL
NPRSONLHAVISALPRAWGLTNRVHGRVINDTFVQFIQSEIDLLSVIRREPWLNNW
FVTAORWEVNLTFHLLTSIELVQMGRIPIILYVCEETALETAHELGLIITLDFDSTT
TQIAYIRVRIRFEGITDRIRFLRIIFDSGETALISFOYERLRRITCCSCFRMTHRNSC
LYRQIESLHRVTNTAQRNVREVFMDENLRSSMNSQSSSESPFPIDPPRIPH
PPLNPDELVAAYIPHTRATSLPNFAGPLPQVPLRRNVYDERSNIQPFSGPFAAHSR
LVEVGESSRQGETQNVHTVEKGDSSKRNMGPKPKDDARKSNEDEHMGGLIKPKPK
KR"
complement(14018..19305)
/gene="At2g13460"
/note="T26C18.5"
/pseudo
complement(16253..16622)
/note="DNA sequence related to non-LTR Tall family of
retrotransposons"
complement(19847..19950)
/rpt_family="(TAA)n"
23014..23118
/rpt_family="(TAAA)n"
23286..>23486
/note="Sequence from clone T10F5"
BASE COUNT 7992 a 4316 c 3907 g 7271 t
ORIGIN

Query Match 77.4%; Score 17.8; DB 8; Length 23486;
Best Local Similarity 90.5%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aaatccatgtaaagcagcagg 22
|||||
Db 14763 AAATCCATGTAAGCAGATCG 14783

Search completed: February 15, 2002, 18:55:04
Job time: 20543 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:01:59 ; Search time 868.33 seconds  
(without alignments)  
22.709 Million cell updates/sec

Title: US-09-698-903B-11  
Perfect score: 23  
Sequence: 1 gaattccatgtaagcagcaggg 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	23	100.0	23	22	AA07000	PCR primer B01, to
2	23	100.0	415	22	AD06997	Right (5') border
3	23	100.0	416	22	AD06999	Left (3') border f
4	18.8	81.7	1152	21	AC46330	Arabidopsis thalia
5	17.4	75.7	1567	21	AC79889	Human secreted pro
6	17.2	74.8	1153	21	AC39525	Arabidopsis thalia
7	17.2	74.8	2857	22	AA18005	Human cDNA sequenc
8	17.2	74.8	4495	22	AA158467	Human polynucleoti
9	17.2	74.8	4775	22	AA160253	Human polynucleoti
10	17.2	74.8	8626	18	AA74989	Saccharomyces cere
11	17.2	74.8	9621	18	AA794548	Mutant YLR087c gen

C 12	16.8	73.0	602	22	AAH29132	Drosophila melanog
C 13	16.8	73.0	1752	22	AAF57408	Human p95 protein
C 14	16.8	73.0	4388	18	AAF60073	Selective marker g
C 15	16.6	72.2	281	14	AAQ0834	Human brain Expres
C 16	16.6	72.2	292	16	AAQ25120	Human gene signatu
C 17	16.6	72.2	292	20	AAV88563	EST clone EY281.
C 18	16.6	72.2	292	21	AAZ46937	Human saccharide-t
C 19	16.6	72.2	549	19	AAV49585	Human stomach canc
C 20	16.6	72.2	552	21	AAA39940	Human TANGO 183 co
C 21	16.6	72.2	575	22	AAF93941	Primer specific fo
C 22	16.6	72.2	692	22	AAH32272	Human ORF factory re
C 23	16.6	72.2	917	21	AAH76546	Human OLFX ORF2101
C 24	16.6	72.2	1026	21	AAH51028	Arabidopsis thalia
C 25	16.6	72.2	1056	21	AAH51420	Arabidopsis thalia
C 26	16.6	72.2	1499	22	AAF93749	Human cDNA encodin
C 27	16.6	72.2	1499	22	AAF29357	Amyloid-beta prote
C 28	16.6	72.2	1504	19	AAV49584	Human stomach canc
C 29	16.6	72.2	1563	22	AAH64828	Human secreted pro
C 30	16.6	72.2	1564	21	AAH37041	Human PRO1248 (UNQ
C 31	16.6	72.2	1564	22	AAF54249	DNA encoding prote
C 32	16.6	72.2	1565	21	AAA39939	Human TANGO 183 CD
C 33	16.6	72.2	1571	20	AAH41983	Human endometrium
C 34	16.6	72.2	1645	21	AAF21831	Human breast and o
C 35	16.6	72.2	1821	21	AAH51418	Arabidopsis thalia
C 36	16.6	72.2	1851	21	AAH51029	Arabidopsis thalia
C 37	16.6	72.2	1929	22	AAH85075	Atherosclerosis-as
C 38	16.6	72.2	1962	22	AAF33128	Human secreted pro
C 39	16.6	72.2	2370	22	AAF26956	Human cancer assoc
C 40	16.6	72.2	2380	21	AAH76858	Human ORFX ORF2413
C 41	16.6	72.2	2383	22	AAH22608	Human cDNA encodin
C 42	16.6	72.2	2393	22	AAH98580	Human EST-derived
C 43	16.6	72.2	2393	22	AAH22844	Human cDNA encodin
C 44	16.6	72.2	2402	21	AAH33327	Human secreted pro
C 45	16.6	72.2	2460	20	AAH58671	Human organic cati

ALIGNMENTS

RESULT 1	
AD07000	
ID	AD07000 standard; DNA; 23 BP.
XX	
XX	AD07000;
XX	
DT	06-AUG-2001 (first entry)
XX	
DE	PCR primer B01, to recognise foreign DNA and flanking sequence of MS-B2.
XX	
XX	MS-B2 elite event; transgenic Brassica plant; transformation event;
KW	male-sterility gene; PCR primer; ss.
XX	
OS	Unidentified.
XX	
PN	WO200131042-A2.
XX	
PD	03-MAY-2001.
XX	
PF	26-OCT-2000; 2000WO-EP10680.
XX	
PR	29-OCT-1999; 99US-0430497.
XX	
PA	(AVET ) AVENTIS CROPS SCIENCE NV.
XX	
PI	Weston B, De Beuckeleer M;
XX	
DR	WPI; 2001-300517/31.
XX	
PT	Transgenic Brassica plants, seeds, cells or tissues, characterized by
PT	harboring specific transformation events, particularly by presence of
PT	male-sterility gene, at specific location in its genome -
XX	
PS	Claim 1; Page 33; 53pp; English.

XX CC The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.  
 XX CC Sequence 23 BP; 9 A; 4 C; 7 G; 3 T; 0 other;

Query Match 100.0%; Score 23; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.088;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgagg 23  
 |||||  
 Db 1 gaaatccatgtaaacgacgagg 23

RESULT 2  
 AAD06997

ID AAD06997 standard; DNA; 415 BP.

XX AC AAD06997;

XX DT 06-AUG-2001 (first entry)

XX DE Right (5') border flanking region of elite event MS-B2.

XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; ds.

XX OS Chimeric - Agrobacterium sp.

XX OS Chimeric - Brassica sp.

XX FH Key Location/Qualifiers

FT misc\_feature 1..234

FT /\*tag= a  
 /note= "Corresponds to plant DNA"

FT misc\_feature 235..415

FT /\*tag= b

FT /note= "Corresponds to T-DNA"

XX WO200131042-A2.

XX PD 03-MAY-2001.

XX PF 26-OCT-2000; 2000WO-EP10680.

XX PR 29-OCT-1999; 99US-0430497.

XX PA (AVET ) AVENTIS CROPS SCIENCE NV.

XX PI Weston B, De Beuckeleer M;

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -

XX Claim 11; Page 51; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the

CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is right (5') border flanking region of elite event  
 CC MS-B2.

SQ Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 100.0%; Score 23; DB 22; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgagg 23  
 |||||  
 Db 152 gaaatccatgtaaacgacgagg 174

RESULT 3  
 AAD06999/c

ID AAD06999 standard; DNA; 416 BP.

XX AC AAD06999;

XX DT 06-AUG-2001 (first entry)

XX DE Left (3') border flanking region of elite event MS-B2.

XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; ds.

XX OS Chimeric - Agrobacterium sp.

XX OS Chimeric - Brassica sp.

XX FH Key Location/Qualifiers

FT misc\_feature 1..193

FT /\*tag= a

FT /note= "Corresponds to T-DNA"

FT misc\_feature 194..416

FT /\*tag= b

FT /note= "Corresponds to plant DNA"

XX WO200131042-A2.

XX PD 03-MAY-2001.

XX PF 26-OCT-2000; 2000WO-EP10680.

XX PR 29-OCT-1999; 99US-0430497.

XX PA (AVET ) AVENTIS CROPS SCIENCE NV.

XX PI Weston B, De Beuckeleer M;

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -

XX Claim 11; Page 52; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is left (3') border flanking region of elite event  
 CC MS-B2.

SQ Sequence 416 BP; 137 A; 72 C; 54 G; 152 T; 1 other;

	Query Match	100.0%; Score 23; DB 22; Length 416;
	Best Local Similarity	100.0%; Pred. NO. 0.14; Indels 0; Gaps 0;
	Matches 23; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 gaaatcccatgtaaacgcagcagg 23 	
Db	263 GAATCCCATGTAAAGCAGCAGG 241 	
RESULT 4		
ID AAC46330	AAC46330 standard; DNA; 1152 BP.	
XX AC AAC46330;		
XX XX		
DT DT	18-OCT-2000 (first entry)	
DE DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 49743.	
XX XX		
KW KW	Hybridisation assay; genetic mapping; gene expression control;	
KW KW	protein identification; signal transduction pathway;	
OS OS	metabolic pathway; promoter; termination sequence; ss.	
XX XX		
PN PN	Arabidopsis thaliana.	
PD PD	EP1033405-A2.	
PF PF	06-SEP-2000.	
XX XX	25-FEB-2000; 2000EP-0301439.	
PR PR	25-FEB-1999; 99US-0121825.	
PR PR	05-MAR-1999; 99US-0123180.	
PR PR	09-MAR-1999; 99US-0123548.	
PR PR	23-MAR-1999; 99US-0125788.	
PR PR	25-MAR-1999; 99US-0126264.	
PR PR	29-MAR-1999; 99US-0126785.	
PR PR	01-APR-1999; 99US-0127462.	
PR PR	06-APR-1999; 99US-0128234.	
PR PR	08-APR-1999; 99US-0128714.	
PR PR	16-APR-1999; 99US-0129845.	
PR PR	19-APR-1999; 99US-0130077.	
PR PR	21-APR-1999; 99US-0130449.	
PR PR	23-APR-1999; 99US-0130510.	
PR PR	28-APR-1999; 99US-0130891.	
PR PR	30-APR-1999; 99US-0131449.	
PR PR	30-APR-1999; 99US-0132048.	
PR PR	04-MAY-1999; 99US-0132407.	
PR PR	05-MAY-1999; 99US-0132484.	
PR PR	06-MAY-1999; 99US-0132485.	
PR PR	06-MAY-1999; 99US-0132487.	
PR PR	07-MAY-1999; 99US-0132863.	
PR PR	11-MAY-1999; 99US-0134256.	
PR PR	14-MAY-1999; 99US-0134284.	
PR PR	14-MAY-1999; 99US-0134218.	
PR PR	14-MAY-1999; 99US-0134219.	
PR PR	14-MAY-1999; 99US-0134221.	
PR PR	18-MAY-1999; 99US-0134370.	
PR PR	19-MAY-1999; 99US-0134768.	
PR PR	20-MAY-1999; 99US-0134941.	
PR PR	21-MAY-1999; 99US-0135124.	
PR PR	24-MAY-1999; 99US-0135353.	
PR PR	25-MAY-1999; 99US-0135629.	
PR PR	27-MAY-1999; 99US-0136031.	
PR PR	28-MAY-1999; 99US-0136392.	
PR PR	01-JUN-1999; 99US-0136782.	
PR PR	03-JUN-1999; 99US-0137222.	
PR PR	04-JUN-1999; 99US-0137528.	
PR PR	07-JUN-1999; 99US-0137502.	
PR PR	08-JUN-1999; 99US-0137724.	
PR PR	10-JUN-1999; 99US-0138094.	
PR PR	10-JUN-1999; 99US-0138540.	
PR PR	10-JUN-1999; 99US-0138847.	
PR PR	14-JUN-1999; 99US-0139119.	
PR PR	16-JUN-1999; 99US-0139452.	
PR PR	17-JUN-1999; 99US-0139453.	
PR PR	18-JUN-1999; 99US-0139452.	
PR PR	18-JUN-1999; 99US-0139455.	
PR PR	18-JUN-1999; 99US-0139456.	
PR PR	18-JUN-1999; 99US-0139457.	
PR PR	18-JUN-1999; 99US-0139458.	
PR PR	18-JUN-1999; 99US-0139459.	
PR PR	18-JUN-1999; 99US-0139460.	
PR PR	18-JUN-1999; 99US-0139461.	
PR PR	18-JUN-1999; 99US-0139462.	
PR PR	18-JUN-1999; 99US-0139463.	
PR PR	18-JUN-1999; 99US-0139464.	
PR PR	18-JUN-1999; 99US-0139465.	
PR PR	18-JUN-1999; 99US-0139466.	
PR PR	18-JUN-1999; 99US-0139467.	
PR PR	18-JUN-1999; 99US-0139468.	
PR PR	18-JUN-1999; 99US-0139469.	
PR PR	18-JUN-1999; 99US-0139470.	
PR PR	18-JUN-1999; 99US-0139471.	
PR PR	18-JUN-1999; 99US-0139472.	
PR PR	18-JUN-1999; 99US-0139473.	
PR PR	18-JUN-1999; 99US-0139474.	
PR PR	18-JUN-1999; 99US-0139475.	
PR PR	18-JUN-1999; 99US-0139476.	
PR PR	18-JUN-1999; 99US-0139477.	
PR PR	18-JUN-1999; 99US-0139478.	
PR PR	18-JUN-1999; 99US-0139479.	
PR PR	18-JUN-1999; 99US-0139480.	
PR PR	18-JUN-1999; 99US-0139481.	
PR PR	18-JUN-1999; 99US-0139482.	
PR PR	18-JUN-1999; 99US-0139483.	
PR PR	18-JUN-1999; 99US-0139484.	
PR PR	18-JUN-1999; 99US-0139485.	
PR PR	18-JUN-1999; 99US-0139486.	
PR PR	18-JUN-1999; 99US-0139487.	
PR PR	18-JUN-1999; 99US-0139488.	
PR PR	18-JUN-1999; 99US-0139489.	
PR PR	18-JUN-1999; 99US-01394	

PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 26-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 81.7%; Score 18.8; DB 21; Length 1152;  
 Best Local Similarity 90.9%; Pred. No. 15;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 2 aaatccatgttaagcagcaggg 23  
 |||||||  
 Db 1088 aaatccatgttaagcagcaggg 1109

## RESULT 5

AAC79889

ID AAC79889 standard; cDNA; 1567 BP.

XX AAC79889;

AC AAC79889;

DT 09-FEB-2001 (first entry)

DE Human secreted protein encoding cDNA for gene 41.

XX Human; secreted protein; cytostatic; antiarthritic; antiasthmatic;  
 immunosuppressive; antiarteriosclerotic; antiinflammatory; nootropic;  
 neuroprotective; antidiabetic; tranquiliser; vulnerary; antibacterial;  
 antipsoriatic; antiarrhythmic; antirheumatic; cardiant; anti-HIV;  
 autoimmune disorder; allergic condition; cardiovascular disorder;  
 cancer; neurological disease; tissue repair; ss.

OS Homo sapiens.

XX WO200055176-A2.

PN 21-SEP-2000.

PD 09-MAR-2000; 2000WO-US06057.

PF 12-MAR-1999; 99US-0124142.

PR 11-JUN-1999; 99US-0138597.

PR 03-DEC-1999; 99US-0168666.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-638176/61.

DR P-PSDB; RAB44870.

XX Novel 49 human secreted proteins useful for diagnosis, prevention and  
 treatment of disorders including neurological, cell proliferative,  
 cardiovascular, and autoimmune/inflammatory disorders and microbial  
 infections -

Claim 1a; Page 356-357; 405pp; English.

XX This invention describes a novel isolated polypeptide (I) comprising an  
 amino acid sequence at least 95 % identical to a polypeptide sequence  
 selected from 49 polypeptides encoded by polynucleotide sequences  
 included in American Type Culture Collection (ATCC) deposit number  
 203917, defined in the specification. The products of the invention have  
 cytostatic, antiarthritic, antiasthmatic, immunosuppressive, nootropic,  
 antiarteriosclerotic, antiinflammatory, neuroprotective, antidiabetic,  
 tranquiliser, vulnerary, antibacterial, antipsoriatic, antiarrhythmic,  
 antirheumatic, cardiant and anti-HIV activity. (I) or a nucleic acid (II)  
 encoding (I) is useful for preventing, treating or ameliorating a medical  
 condition and for diagnosing a pathological condition or susceptibility  
 to the condition. (I) is useful for identifying a binding partner which  
 affects the activity of the polypeptide and for identifying an activity  
 in a biological sample. (I), (II) or an antibody (IV) specific to (I) is  
 also useful for treating or preventing a disease, disorder or condition  
 associated with aberrant expression of (I). Diseases treated or condition  
 include immune disorders such as autoimmune diseases, blood protein  
 disorders, anemia, allergic reactions and conditions such as asthma,  
 organ rejection or graft-versus-host disease, inflammation, hyper  
 proliferative disorders, cardiovascular disorders such as arterioarterial  
 fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ  
 regeneration, cancer, neovascular glaucoma, diabetic retinopathy,  
 rheumatoid arthritis, psoriasis, diseases associated with increased  
 apoptosis that include acquired immunodeficiency syndrome (AIDS),  
 neurological diseases such as Parkinson's disease, viral, bacterial,  
 fungal or parasitic diseases. They are also used to repair, replace or  
 protect tissue damage by congenital defects, to treat trauma, in surgery,  
 including cosmetic plastic surgery, to treat fibrosis, reperfusion injury

CC or systemic cytokine damage, to stimulate chondrocyte growth, to prevent  
CC skin aging due to sunburn, to change a mammal's mental state or physical  
CC state by influencing biorhythms, cardiac rhythms, depression, memory,  
CC stress and for accelerating wound healing. (I), (II) and/or their agonist  
CC or antagonist are useful as food additives or preservatives to increase  
CC or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamin, mineral or other nutritional components. (I) is  
CC useful for screening therapeutic compounds. (II) is useful in forensic  
CC biology for detecting DNA sequences and as diagnostic probes for  
CC detecting the presence of specific mRNA in a particular cell type.  
XX  
SQ Sequence 1567 BP; 303 A; 510 C; 463 G; 291 T; 0 other;

Query Match 75.7%; Score 17.4; DB 21; Length 1567;

Best Local Similarity 94.7%; Pred. No. 71;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 atccatgtaagcagcagg 22  
||||||| |||||||

Db 805 atccatgtaagcagcagg 823

#### RESULT 6

AAC39525  
ID AAC39525 standard; DNA: 1153 BP.

XX AC AAC39525;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 24952.

XX KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0126785.

PR 06-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130510.

PR 30-APR-1999; 99US-0130891.

PR 04-MAY-1999; 99US-0131449.

PR 05-MAY-1999; 99US-0132048.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.

```
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 74.8%; Score 17.2; DB 21; Length 1153;
Best Local Similarity 86.4%; Pred. No. 84;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaaagcagcagg 23
|||||||
Db 1088 aaatccatgtaaatacagtaagg 1109

RESULT 7
AAH18005
ID AAH18005 standard; cDNA; 2857 bp.
XX
AC AAH18005;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17809.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;
XX
WPI; 2001-318749/34.
XX
PR Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
XX
PS Claim 8; SEQ ID 17809; 2537pp + CD ROM; English.
XX
The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination
of the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesizing polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
```







```

Db 7927 AAATTAATGTAAGCACCAGG 7906
||||| ||||||||| |||||
RESULT 12
AAH29132/c
ID AAH29132 standard; DNA; 602 BP.
XX
AC AAH29132;
XX
DT 17-JUL-2001 (first entry)
XX
DE Drosophila melanogaster essential gene fragment, SEQ ID NO: 321.
XX
KW Drosophila melanogaster; fruit fly; essential gene; screening assay;
KW pesticide; crop protection; chromosome 3; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200118547-A1.
XX
PD 15-MAR-2001.
XX
PF 06-SEP-2000; 2000WO-GB03444.
XX
PR 07-SEP-1999; 99GB-0021009.
XX
PA (UNIU ) UNIV GLASGOW.
XX
PI Davies RW, Kaiser K, Yang MY;
XX
DR WPI; 2001-281436/29.
XX
PT Screening assays for used for identifying compounds having a
PT physiological effect on proteins identified as being essential -
XX
PS Claim 1; Page 340-341; 695pp; English.
XX
CC The present sequence is part of an essential gene from Drosophila
CC melanogaster. Lack of expression of the protein encoded by this
CC gene leads to a lethal or semi-lethal phenotype. The invention
CC relates to 902 nucleic acid sequences from genes encoding proteins
CC which are thought to be essential, and to a screening assay for
CC identifying compounds which have a physiological effect on these
CC proteins. Suitable compounds are useful as pesticides and may be used
CC in conjunction with other pesticides and herbicides for crop
CC protection. The gene corresponding to the present sequence is located
CC on chromosome 3.
XX
SQ Sequence 602 BP; 159 A; 131 C; 120 G; 192 T; 0 other;

Query Match 73.0%; Score 16.8; DB 22; Length 602;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacagcag 21
| ||||| ||||| |||||
Db 409 AGATCCATGTAATCAGCAG 390

RESULT 13
AAF57408/c
ID AAF57408 standard; DNA; 1752 BP.
XX
AC AAF57408;
XX
DT 11-JUN-2001 (first entry)
XX
DE Human p95 protein encoding DNA.
XX
KW Rab; Rho; GTPase; pharmaceutical; cancer; anti-HIV; tuberculostatic;
KW protozoacide; atidiabetic; nootropic; neuroprotective; dermatological;

```

```

KW antipsoriatic; antiinflammatory; antiallergic; antipyretic; cytostatic;
KW antibacterial; gynecological; antiatherosclerotic; gene therapy; human;
KW p100; p95; p60; p45; p25; endocytic trafficking; GTPase effector; ds.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1752
FT /*tag= a
FT /product= "p95"
XX
PN WO200120022-A1.
XX
PD 22-MAR-2001.
XX
PF 18-SEP-2000; 2000WO-EP09130.
XX
PR 16-SEP-1999; 99EP-0118385.
XX
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Nielsen E, Christophoridis S, Murphy C, Zerial M, De Renzis S;
XX
DR WPI; 2001-257888/26.
DR P-PSDB; AAB62180.
XX
PT Use of effectors of GTPase as target in a in vitro/vivo assay for
PT detecting substances for prophylaxis, treatment of cancer, cell
PT migration disorders, e.g. Alzheimer's, infectious diseases, diabetes,
PT atherosclerosis -
XX
PS Claim 10; Page 70-71; 76pp; English.
XX
CC The invention relates to the use of effectors/regulators for Rab and Rho
CC GTPases in in vitro or in vivo assays. The method is useful for detecting
CC substances useful as pharmaceutical agents for the prophylaxis or
CC treatment of cancer and other proliferative, invasive or cell migration
CC disorders such as endometriosis, atherosclerosis, inflammatory and
CC allergic diseases, infectious diseases, diabetes, Alzheimer's disease and
CC skin repair diseases such as psoriasis. The infectious diseases include
CC AIDS, tuberculosis, pseudotuberculosis, cholera, gastroenteritis, enteric
CC fever, malaria, typhus, diseases caused by pathogens such as Listeria,
CC Mycobacterium, Staphylococcus, Toxoplasma, Trypanosoma, Salmonella,
CC Legionella, Leishmania, Coxiella, Shigella, Yersinia, Neisseria, Vibrio,
CC Bartonella. The cancer includes benign tumor, malignant tumor, carcinoma,
CC leukemia, glioma or a neuroblastoma, in particular lung carcinoma,
CC osteosarcoma, lymphoma, breast, bile, intestine, kidney, ovary, stomach,
CC brain, prostate, liver and every tumor that invades other tissues and
CC organs distinct from its site of origin. The assay is highly sensitive
CC and advantageous in the selectivity of the targets. The present sequence
CC represents the DNA encoding a human p95 protein. A multiprotein complex
CC including p100, p95, p60, p45, p25 (sequences AAF57404-408) acts as a
CC GTPase effector/regulator and has the function of endocytic trafficking.
XX
SQ Sequence 1752 BP; 496 A; 425 C; 444 G; 387 T; 0 other;

Query Match 73.0%; Score 16.8; DB 22; Length 1752;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaaatccatgtaaacgacagc 20
| ||||| ||||| |||||
Db 712 GAAATCCATGCAGCAGCAGCA 693

RESULT 14
AAT60073
ID AAT60073 standard; DNA; 4388 BP.
XX
AC AAT60073;
XX
DT 13-MAY-1997 (first entry)

```

XX DE Selective marker gene.  
XX KW Selective marker; enzyme; biosynthetic system; leucine synthesis;  
XX KW Rhizomucor pusillus; protein production; ss.  
XX OS Synthetic.  
XX PN JP09009971-A.  
XX PD 14-JAN-1997.  
XX PF 30-JUN-1995; 95JP-0166115.  
XX PR 30-JUN-1995; 95JP-0166115.  
XX PA (MEIT ) MEITO SANGYO KK.  
XX DR WPI; 1997-126428/12.  
XX PT Novel selective marker gene - used in a transformation system of  
XX PT Rhizomucor pusillus for protein production  
XX PS Claim 1; Page 9-10; 12pp; Japanese.  
XX CC This sequence represents the selective marker gene of the invention. This  
XX CC gene encodes an enzyme relating to the biosynthetic system of leucine  
XX CC synthesis. This sequence can be used in a recombinant vector for the  
XX CC transformation of Rhizomucor pusillus. The recombinant vector also  
XX CC includes another DNA sequence encoding a protein. The plasmid and the  
XX CC transformed Rhizomucor pusillus can then be used for the efficient  
XX CC production of useful proteins and peptides, which are encoded by the  
XX CC second DNA sequence in the vector.  
XX SQ Sequence 4388 BP; 1144 A; 1094 C; 1077 G; 1073 T; 0 other;

Query Match 73.0%; Score 16.8; DB 18; Length 4388;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 gaaatccatgtaaaagcagca 20  
Db 611 gaaatgcataaaagcagca 630  
RESULT 15  
AAQ60834  
ID AAQ60834 standard; DNA; 281 BP.  
XX AC AAQ60834;  
XX DT 16-MAR-1994 (first entry)  
XX DE Human brain Expressed Sequence Tag EST00937.  
XX KW Gene transcription product; genetic markers; tagging; in vivo;  
XX KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
XX OS Homo sapiens.  
XX PN WO9316178-A.  
XX PD 19-AUG-1993.  
XX PF 12-FEB-1993; 93WO-US01294.  
XX PR 12-FEB-1992; 92US-0837195.  
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
XX PI Adams MD, Moreno RF, Venter CJ;  
XX XX

DR WPI; 1993-272882/34.  
XX Enriched oligonucleotides and corresp. sequences - used as  
XX PT markers for human genes transcribed in-vivo, facilitate tagging  
XX PT of most human genes  
XX XX  
XX PS Example 4; Page 398; 500pp; English.  
XX CC The Expressed Sequence Tag was isolated from a human brain cDNA  
XX CC library as part of a large set of ESTs which can be used as markers  
XX CC for human genes transcribed in vivo. They can be used to facilitate  
XX CC tagging of most human genes, for mapping locations of expressed genes  
XX CC on chromosomes, for individual or forensic identification, for mapping  
XX CC locations of disease-associated genes, for identification of tissue  
XX CC type, and for prepn. of antisense sequences, probes and constructs.  
XX CC EST00937 has a "poor" coding probability as evaluated using the  
XX CC coding-region prediction program CRM. See also AAQ59041-Q61440.  
XX SQ Sequence 281 BP; 45 A; 81 C; 91 G; 61 T; 3 other;  
Query Match 72.2%; Score 16.6; DB 14; Length 281;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 gaaatccatgtaaaagcagcagg 23  
Db 86 gaagtgaatgcaaaagcagcagg 108  
Search completed: February 15, 2002, 19:02:02  
Job time: 20766 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:09:12 ; Search time 353.79 Seconds  
(without alignments)  
14.723 Million cell updates/sec

Title: US-09-698-903B-11  
Perfect score: 23  
Sequence: 1 gaatccatgtaaacgacgagg 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues  
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.2	74.8	8574	4	US-08-894-344C-1
C 2	17.2	74.8	9521	4	US-09-125-028-1
C 3	16.6	72.2	2460	4	US-08-964-127-1
C 4	16.6	72.2	9751	4	US-09-238-303-7
C 5	16.2	70.4	27	3	US-08-832-399-9
C 6	16.2	70.4	27	4	US-09-372-498-9
C 7	16.2	70.4	717	2	US-08-706-702-25
C 8	16.2	70.4	717	3	US-08-706-702-25
C 9	16.2	70.4	1351	4	US-09-002-114-2
C 10	16.2	70.4	1608	2	US-08-899-371-2
C 11	16.2	70.4	1613	2	US-08-899-371-1
C 12	16.2	70.4	5183	3	US-09-039-555B-18
C 13	16.2	70.4	5243	2	US-08-414-335-2
C 14	16.2	70.4	16442	3	US-08-781-891-208
C 15	15.8	68.7	199	2	US-08-250-346-21
C 16	15.8	68.7	394	2	US-07-885-089B-15
C 17	15.8	68.7	394	2	US-07-885-089B-17
C 18	15.8	68.7	3410	4	US-09-020-955-110
C 19	15.8	68.7	3410	4	US-09-030-607-110
C 20	15.8	68.7	3466	1	US-08-468-036-38
C 21	15.8	68.7	3466	2	US-08-376-843-38
C 22	15.6	67.8	1098	2	US-08-948-616-6
C 23	15.6	67.8	1098	2	US-09-193-510-6
C 24	15.6	67.8	1098	4	US-09-368-402-6
C 25	15.6	67.8	1488	4	US-09-171-969-1
C 26	15.4	67.0	2266	2	US-09-213-767-1
C 27	15.4	67.0	152331	3	US-09-128-155-16

C 28	15.4	67.0	176373	3	US-09-128-155-17
C 29	15.2	66.1	681	5	PCT-US94-04361-22
C 30	15.2	66.1	1338	2	US-08-484-993B-7
C 31	15.2	66.1	1338	2	US-08-484-158B-7
C 32	15.2	66.1	1338	2	US-08-484-596A-7
C 33	15.2	66.1	1338	2	US-08-480-150A-7
C 34	15.2	66.1	1338	3	US-08-458-731-7
C 35	15.2	66.1	1338	3	US-08-149-223A-7
C 36	15.2	66.1	2975	1	US-08-368-281-1
C 37	15.2	66.1	5107	4	US-08-910-647-3
C 38	15.2	66.1	18994	1	US-08-459-586-4
C 39	15.2	66.1	18994	2	US-08-282-696-4
C 40	15	65.2	35	2	US-08-094-534-18
C 41	15	65.2	35	2	US-08-581-543-18
C 42	15	65.2	35	5	PCT-US94-08000-18
C 43	15	65.2	39	1	US-08-094-534-17
C 44	15	65.2	39	2	US-08-581-543-17
C 45	15	65.2	39	5	PCT-US94-08000-17

ALIGNMENTS

RESULT 1  
US-08-894-344C-1/c  
; Sequence 1, Application US/08894344C  
; Patent No. 6172196  
; GENERAL INFORMATION:  
; APPLICANT: KAWASAKI, Hideki  
; APPLICANT: TOKAI, Masaya  
; APPLICANT: KIKUCHI, Yasuhiro  
; APPLICANT: OUCHI, Kozo  
; TITLE OF INVENTION: DNA ENCODING PROTEIN COMPLEMENTING  
; TITLE OF INVENTION: YEAST  
; TITLE OF INVENTION: LOW TEMPERATURE-SENSITIVE FERMENTABILITY  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: FITZPATRICK, CELLA, HARPER & SCINTO  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10112-3801

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette - 3.50 inch, 1440 kb storage.  
COMPUTER: IBM PS/V  
OPERATING SYSTEM: MS-DOS Ver3.30  
SOFTWARE: PATENT AID Ver1.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,344C  
FILING DATE: 15-AUGUST-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP343700/95  
FILING DATE: 28-DECEMBER-1995  
APPLICATION NUMBER: PCT/JP96/03862  
FILING DATE: 27-DECEMBER-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Perry, Lawrence S.  
REGISTRATION NUMBER: 31865  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-218-2100  
TELEFAX: 212-218-2200  
INFORMATION FOR SEQ ID NO: 1 :  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8874 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Saccharomyces cerevisiae  
STRAIN: X2180-1B  
FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 1 to 8874
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 1291 to 1296
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 4388 to 4393
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 5927 to 5032
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 7675 to 7680
; IDENTIFICATION METHOD: S
US-08-894-344C-1

```

```

Query Match 74.8%; Score 17.2; DB 4; Length 8874;
Best Local Similarity 86.4%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 2 aaatccatgtaaagcagcagg 23
||||| ||||| ||||| |||||
Db 7552 AAATTAATGTAAGCACCAGG 7531

```

```

RESULT 2
US-09-125-028-1/c
; Sequence 1, Application US/09125028A
; Patent No. 6190707
; GENERAL INFORMATION:
; APPLICANT: WADOUX, Isabelle
; APPLICANT: COLAVIZZA, Didier
; APPLICANT: LOIEZ, Annie
; TITLE OF INVENTION: NOVEL COLD-SENSITIVE BREAD-MAKING YEASTS
; FILE REFERENCE: levure sensible froid
; CURRENT APPLICATION NUMBER: US/09/125.028A
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: PCT/FR97/00254
; EARLIER FILING DATE: 1997-02-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 9621
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9318)
; OTHER INFORMATION: Use of n signifies any of g, a, c or t
US-09-125-028-1

```

```

Query Match 74.8%; Score 17.2; DB 4; Length 9621;
Best Local Similarity 86.4%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 2 aaatccatgtaaagcagcagg 23
||||| ||||| ||||| |||||
Db 7927 AAATTAATGTAAGCACCAGG 7906

```

```

RESULT 3
US-08-964-127-1/c
; Sequence 1, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John

```

```

; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,127
; FILING DATE: 06-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crews, Ph.D., L. Lee
; REGISTRATION NUMBER: P-43,567
; REFERENCE/DOCKET NUMBER: 07334/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 498...2057
US-08-964-127-1

```

```

Query Match 72.2%; Score 16.6; DB 4; Length 2460;
Best Local Similarity 82.6%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 gaaatccatgtaaagcagcagg 23
||||| ||||| ||||| |||||
Db 2242 GAAGTGAATGCAAGCAGCAGG 2220

```

```

RESULT 4
US-09-238-303-7
; Sequence 7, Application US/09238303B
; Patent No. 6284253
; GENERAL INFORMATION:
; APPLICANT: Barr, Margaret C.
; TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequence
; FILE REFERENCE: 18617.0059
; CURRENT APPLICATION NUMBER: US/09/238,303B
; CURRENT FILING DATE: 1999-01-28
; EARLIER APPLICATION NUMBER: US 60/072,927
; EARLIER FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 7
; LENGTH: 9751
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of
; OTHER INFORMATION: a Pallas's cat feline immunodeficiency virus
US-09-238-303-7

```

```

; TITLE OF INVENTION: AND SPICE VARIANTS THEROF
; FILE REFERENCE: GH-50020-1
; CURRENT APPLICATION NUMBER: US/09/372,498
; PRIOR FILING DATE: 1999-08-11
; CURRENT APPLICATION NUMBER: 08/832,399
; PRIOR FILING DATE: 1997-04-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 27
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-372-498-9

Query Match 70.4%; Score 16.2; DB 4; Length 27;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 3; Indels

QY 1 gaatccatgttaaagcagcag 21
   ||| ||||| ||| |||||
Db 6 gaattccatggaaccagcag 26

RESULT 7
US-08-706-702-25/c
; Sequence 25, Application US/08706702
; Patent No. 5948614
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Deb K.
; TITLE OF INVENTION: Cloned DNA Polymerases from Thermotoga
; TITLE OF INVENTION: Maritima and Mutants Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,702
; FILING DATE: 06-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/689,807
; FILING DATE: 14-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,400
; FILING DATE: 02-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,759
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,397
; FILING DATE: 02-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/525,057
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0942.2800006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:

```

;  
; LENGTH: 717 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: DNA (genomic)  
US-08-706-702-25

Query Match 70.4%; Score 16.2; DB 2; Length 717;  
Best Local Similarity 78.3%; Pred. No. 50;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcagg 23  
| |||| ||||| |||||  
Db 673 GGNATCCNGGTAAAGCGCAGG 651

RESULT 8  
US-08-706-706-25/c  
; Sequence 25, Application US/08706706  
; Patent No. 6015668  
; GENERAL INFORMATION:  
; APPLICANT: Hughes, A. John  
; APPLICANT: Chatterjee, Deb K.  
; TITLE OF INVENTION: Cloned DNA Polymerases from Thermotoga and  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 New York Ave., N.W., Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08706,706  
; FILING DATE: 06-SEP-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/689,807  
; FILING DATE: 14-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/537,400  
; FILING DATE: 02-OCT-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/370,190  
; FILING DATE: 01-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/316,423  
; FILING DATE: 30-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/576,759  
; FILING DATE: 21-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/537,397  
; FILING DATE: 02-OCT-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/525,057  
; FILING DATE: 08-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REGISTRATION NUMBER: 32,893  
; REFERENCE/DOCKET NUMBER: 0942.2800005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 717 base pairs

;  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: DNA (genomic)  
US-08-706-706-25

Query Match 70.4%; Score 16.2; DB 3; Length 717;  
Best Local Similarity 78.3%; Pred. No. 50;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcagg 23  
| |||| ||||| |||||  
Db 673 GGNATCCNGGTAAAGCGCAGG 651

RESULT 9  
US-09-002-114-2/c  
; Sequence 2, Application US/09002114  
; Patent No. 6274720  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Shah, Purvi  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN PREPRONEUROTENSIN/NEUROMEDIN N  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/002,114  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0450 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1351 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-002-114-2

Query Match 70.4%; Score 16.2; DB 4; Length 1351;  
Best Local Similarity 85.7%; Pred. No. 57;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaaagcagcagg 22  
||| |||| ||||| |||||  
Db 384 AAGCCATCTAAAGCAGTAGG 364

RESULT 10  
US-08-899-371-2



```
; Sequence 2, Application US/08899371
; Patent No. 5962227
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Ronald P.
; APPLICANT: Andree, Karl B.
; APPLICANT: Antonio, Dolores B.
; TITLE OF INVENTION: A DNA-Based Diagnostic Test for
; TITLE OF INVENTION: Detecting Myxobolus, the Cause of Salmonid Whirling
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsens and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,371
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 60/022,734
; FILING DATE: 26-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-081310US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..1608
; OTHER INFORMATION: /note= "18S rRNA gene of Myxobolus
; insidiosus"
;
US-08-899-371-2

Query Match 70.4%; Score 16.2; DB 2; Length 1608;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaaagcagcagg 22
| | | | | | | | | | | | | | | |
Db 312 ACATCCATGGAGGCAGCAGG 332

RESULT 11
US-08-899-371-1
; Sequence 1, Application US/08899371
; Patent No. 5962227
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Ronald P.
; APPLICANT: Andree, Karl B.
; APPLICANT: Antonio, Dolores B.
; TITLE OF INVENTION: A DNA-Based Diagnostic Test for
; TITLE OF INVENTION: Detecting Myxobolus, the Cause of Salmonid Whirling
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Townsens and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,371
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/022,734
; FILING DATE: 26-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-081310US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..1613
; OTHER INFORMATION: /note= "18S rRNA gene of Myxobolus
; cerebrealis"
;
US-08-899-371-1

Query Match 70.4%; Score 16.2; DB 2; Length 1613;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaaagcagcagg 22
| | | | | | | | | | | | | | | |
Db 315 ACATCCATGGAGGCAGCAGG 335

RESULT 12
US-09-039-555B-18/c
; Sequence 18, Application US/09039555B
; Patent No. 6033856
; GENERAL INFORMATION:
; APPLICANT: Koerner, Kathrin
; APPLICANT: Mueller, Rolf
; APPLICANT: Sadlacek, Hans-Harald
; TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/09/039,555B  
FILING DATE: 16-MAR-1998  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 19710643.9  
FILING DATE: 14-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016779/0131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5183 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-039-555B-18

Query Match 70.4%; Score 16.2; DB 3; Length 5183;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaaagcagcagg 22  
||||||| |||||||||  
Db 741 AAATCCAGCTATAGCAGCAGG 721

RESULT 13  
US-08-414-335-2  
Sequence 2, Application US/08414335  
Patent No. 5907078  
GENERAL INFORMATION:  
APPLICANT: GREENBERT, No. 5907078man M  
APPLICANT: MATUSIK, Robert J  
APPLICANT: ROSEN, Jeffrey M  
TITLE OF INVENTION: TRANSGENIC MOUSE MODEL FOR PROSTATE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare Ltd.  
STREET: 1203 Crystal Plaza Bldg. I, 2001 Jefferson  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/414,335  
FILING DATE: 31-MAR-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: FALLOW, Charles W  
REGISTRATION NUMBER: 28,946  
REFERENCE/DOCKET NUMBER: 1027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5243 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

US-08-414-335-2  
TOPOLOGY: linear

Query Match 70.4%; Score 16.2; DB 2; Length 5243;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaaagcagcagg 22  
||||||| |||||||||  
Db 4443 AAATCCAGCTATAGCAGCAGG 4463

RESULT 14  
US-08-781-891-208/C  
Sequence 208, Application US/08781891  
Patent No. 6090620  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Yu, Chang-En  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
TITLE OF INVENTION: WERNER'S SYNDROME  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16442 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-208

Query Match 70.4%; Score 16.2; DB 3; Length 16442;  
Best Local Similarity 85.7%; Pred. No. 94;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcag 21  
||||||| |||||||||  
Db 12904 GAAATCCAGTATAAGTCACAG 12884

RESULT 15  
US-08-250-346-21/c  
Sequence 21, Application US/08250346  
Patent No. 5939255  
GENERAL INFORMATION:  
APPLICANT: ANAND, Rakesh

APPLICANT: MARKHAM, Alexander F  
APPLICANT: SMITH, John C  
APPLICANT: ANWAR, Rashida  
APPLICANT: RILEY, John H  
APPLICANT: OGILVIE, Donald J  
APPLICANT: ELVIN, Paul  
TITLE OF INVENTION: DIAGNOSTIC METHOD  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSER: CUSHMAN DABBY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/250,346  
FILING DATE: 27-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/115,675  
FILING DATE: 02-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/879,117  
FILING DATE: 04-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/578,616  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 8920211.3  
FILING DATE: 07-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: BIRD, Donald J.  
REGISTRATION NUMBER: 25,323  
REFERENCE/DOCKET NUMBER: 205356/PHM35417/USC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 199 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-250-346-21

Query Match 68.7%; Score 15.8; DB 2; Length 199;  
Best Local Similarity 89.5%; Pred. No. 60;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 tccatgtaaacgacgaggg 23  
|||||  
Db 178 TCCAGGAAAGACGACGGG 160

Search completed: February 15, 2002, 19:09:16  
Job time: 14730 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:04:13 ; Search time 9904.61 seconds  
(without alignments)  
24.953 Million cell updates/sec

Title: US-09-698-903B-11  
Perfect score: 23  
Sequence: 1 gaaatccatgtaagcagcagg 23  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.8	81.7	619	FR0032239	AL028607 Fugu rubr
2	18.8	81.7	916	CNS050PT	AL315856 Tetraodon
3	18.8	81.7	1007	CNS05NNC	AL345585 Tetraodon
c 4	18.4	80.0	342	CNS042531	AZ842531 2M0141J03
c 5	18.4	80.0	522	CNS042531	AQ972989 RPCI-23-3
c 6	18.4	80.0	618	CNS044498	AL030865 Fugu rubr
c 7	18.2	79.1	205	AV119745	AV119745 AV119745
c 8	18.2	79.1	287	AV119723	AV119723 AV119723
c 9	18.2	79.1	444	A1243356	A1243356 qh30g02.x
c 10	18.2	79.1	455	A1243356	AZ628185 1M0460H08
c 11	18.2	79.1	481	BE019857	BE019857 bb60g05.y
c 12	18.2	79.1	533	A2484979	AZ484979 1M0311P16

c 13	18.2	79.1	750	13	AZ750976	RPCI-24-1
c 14	18.2	79.1	809	11	BF027105	601670224
c 15	18.2	79.1	875	11	BG118586	602348330
c 16	18.2	79.1	951	13	CNS0528F	Tetraodon
c 17	17.8	77.4	193	13	AZ658541	1M0535F14
c 18	17.8	77.4	287	10	BE579037	BE579037 Kd53C12.Y
c 19	17.8	77.4	525	10	AV395870	AV395870 AV395870
c 20	17.8	77.4	563	10	AV390896	AV390896 AV390896
c 21	17.8	77.4	564	10	AV387535	AV387535 AV387535
c 22	17.8	77.4	582	13	AZ107460	RPCI-23-3
c 23	17.8	77.4	595	10	AA394343	26024 Lam
c 24	17.8	77.4	618	10	AW661616	833009D12
c 25	17.8	77.4	618	13	AZ558735	RPCI-23-2
c 26	17.8	77.4	626	10	BE337492	894046C11
c 27	17.8	77.4	630	10	BE238191	894046G04
c 28	17.8	77.4	657	10	AW054024	L30-1883T
c 29	17.8	77.4	700	11	BG846208	1024012F1
c 30	17.8	77.4	723	11	BG845360	1024010A0
c 31	17.8	77.4	772	10	AU080746	AU080746 AU080746
c 32	17.4	75.7	221	10	AL507374	AL507374 AL507374
c 33	17.4	75.7	364	10	BE067433	MR4-BT035
c 34	17.4	75.7	604	13	AQ372694	RPCI11-14
c 35	17.4	75.7	857	13	CNS06GUP	AL398183 T7 end of
c 36	17.4	75.7	975	11	BF348682	602020489
c 37	17.4	75.7	1011	13	CNS020BG	AL175381 Tetraodon
c 38	17.2	74.8	158	11	BI268414	BI268414 NF023H01G
c 39	17.2	74.8	226	10	AA298276	EST11389
c 40	17.2	74.8	271	10	AW357388	AW357388 40822 MAR
c 41	17.2	74.8	312	10	AI905219	AI905219 RC-BT078-
c 42	17.2	74.8	312	11	H20500	Yn63G06.si
c 43	17.2	74.8	332	10	BB251925	BB251925 BB251925
c 44	17.2	74.8	340	10	AA012198	TgESTzz16
c 45	17.2	74.8	357	11	W63225	TgESTzy68c1

ALIGNMENTS

RESULT 1

FR0032239 619 bp DNA GSS 27-JUN-1998  
Fugu rubripes GSS sequence, clone 137L16aC10, genomic survey  
sequence.  
AL028607  
AL028607.1 GI:3270721  
GSS: genome survey sequence.  
Takifugu rubripes.  
Takifugu rubripes  
Takifugu rubripes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Takifugu.

REFERENCE  
1 (bases 1 to 619)  
Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,  
Williams,G. and Brenner,S.

TITLE  
Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource  
Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:  
biohelp@hmp.mrc.ac.uk

COMMENT  
Vector: pBluescript II KS  
V.type: phagemid  
PRIMER: KS

DESCR:  
One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.

FEATURES  
source  
1. 619  
/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid 137L16"  
/clone="137L16aC10"

BASE COUNT 131 a 178 c 145 g 116 t 49 others  
ORIGIN

```

Query Match      81.7%; Score 18.8; DB 13; Length 619;
Best Local Similarity 90.9%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aaatccatgtaaacgacgagg 23
| ||||| ||||| ||||| |||||
Db 180 ACATCCAGGTTAAAGCAGCAGG 201

RESULT 2
CNS050PJ      916 bp      DNA      GSS      26-JUL-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 001I09 of library B from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL315856.1 GI:9548744
VERSION AL315856.1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 916)
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Barnot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
2 (bases 1 to 916)
Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
Barnot,A., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
Fischer,C., and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
3 (bases 1 to 916)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
Location/Qualifiers
1. .916
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="001I09"
/clone_lib="B"
/note="Genoscope sequence ID : COAB001AE05C1-end : T7"

BASE COUNT 253 a 223 c 206 g 221 t 13 others
ORIGIN

Query Match      81.7%; Score 18.8; DB 13; Length 916;
Best Local Similarity 90.9%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgagg 22
| ||||| ||||| ||||| |||||
Db 838 GAAATCGATTAAAGCAGCAGG 859

RESULT 3
CNS05NNC      1007 bp      DNA      GSS      26-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 001I09 of library B from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL345585.1 GI:8239355
VERSION AL345585.1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1007)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 1007)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Barnot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1007)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
Location/Qualifiers
1. .1007
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="001I09"
/clone_lib="B"
/note="Genoscope sequence ID : COTB001AE05C1-end : T7"

BASE COUNT 283 a 239 c 217 g 242 t 26 others
ORIGIN

Query Match      81.7%; Score 18.8; DB 13; Length 1007;
Best Local Similarity 90.9%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgagg 22
| ||||| ||||| ||||| |||||
Db 839 GAAATCGATTAAAGCAGCAGG 860

RESULT 4
AZ842531/c      342 bp      DNA      GSS      20-FEB-2001
LOCUS 2M0141J03F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
DEFINITION clone UUGC2M0141J03 F, DNA sequence.
ACCESSION AZ842531
VERSION AZ842531
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 342)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

```



FEATURES  
source

Location/Qualifiers  
1. .618  
/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid 199F09"  
/clone="199F09af9"  
BASE COUNT 109 a 183 c 154 g 141 t 31 others  
ORIGIN

Query Match 80.0%; Score 18.4; DB 13; Length 618;  
Best Local Similarity 86.4%; Pred. No. 4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaagcagcagg 23  
||||| ||||| ||||| ||||| |||||  
Db 362 ACATCCNNGTAAAGCAGCAGG 383

RESULT 7  
LOCUS

AV119745 205 bp mRNA EST 30-JUN-1999  
DEFINITION AV119745 Mus musculus C57BL/6J 10-day embryo Mus musculus cDNA  
ACCESSION AV119745  
VERSION AV119745  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 205)

## AUTHORS

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara,  
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,  
Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,  
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara,  
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,  
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,  
Muranatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE  
JOURNAL

Unpublished (1999)  
Contact: Chie Owa  
Genome Science Laboratory  
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp  
Thermotabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length cDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.

FEATURES  
source

Location/Qualifiers  
1. .205  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="2610305L08"  
/clone\_lib="Mus musculus C57BL/6J 10-day embryo"  
/sex="mixed"  
/dev\_stage="10-day embryo"  
BASE COUNT 42 a 36 c 37 g 90 t  
ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 205;  
Best Local Similarity 87.0%; Pred. No. 4.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaaatccatgtaaagcagcagg 23  
||||| ||||| ||||| ||||| |||||  
Db 192 GAAATCCAAGTAAACACAGAGG 170

RESULT 8  
LOCUS

AV119723 287 bp mRNA EST 30-JUN-1999  
DEFINITION AV119723 Mus musculus C57BL/6J 10-day embryo Mus musculus cDNA  
ACCESSION AV119723  
VERSION AV119723  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 287)

## REFERENCE

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara,  
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,  
Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,  
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara,  
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,  
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,  
Muranatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE  
JOURNAL

Unpublished (1999)  
Contact: Chie Owa  
Genome Science Laboratory  
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp  
Thermotabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length cDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.

FEATURES  
source

Location/Qualifiers  
1. .287  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="2610305J08"  
/clone\_lib="Mus musculus C57BL/6J 10-day embryo"  
/sex="mixed"  
/dev\_stage="10-day embryo"  
BASE COUNT 66 a 57 c 52 g 112 t  
ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 287;  
Best Local Similarity 87.0%; Pred. No. 4.6e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaaatccatgtaaagcagcagg 23  
||||| ||||| ||||| ||||| |||||  
Db 274 GAAATCCAAGTAAACACAGAGG 252

RESULT 9  
LOCUS

AI243356/c 444 bp mRNA EST 01-DEC-1998  
DEFINITION qh30902.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:1846226 3', mRNA sequence.

ACCESSION AI243356  
VERSION AI243356.1  
KEYWORDS EST.  
SOURCE human.



ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 444)  
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1011 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 424.

FEATURES  
source  
1. .444  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1846226"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung MbHL19W, testis NHT, and B-cell NCI-CCAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo." 1 others  
BASE COUNT 138 a 89 c 97 g 119 t 1 others  
ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 444;  
Best Local Similarity 87.0%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgaggg 23  
|||||  
Db 383 GCAACCAATGTAAGCAGCAGGG 361

RESULT 10  
AZ628185 455 bp DNA GSS 13-DEC-2000  
LOCUS IM0480H08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC1M0480H08 F, DNA sequence.  
ACCESSION AZ628185  
VERSION AZ628185.1 GI:11750375  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 455)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606

Fax: 801 585 7177  
Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0480 row: H column: 08  
Seq primer: CGTTGTAAACGACGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 455.

FEATURES  
source  
1. .455  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0480H08"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g114732114gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
BASE COUNT 106 a 114 c 125 g 110 t  
ORIGIN

Query Match 79.1%; Score 18.2; DB 13; Length 455;  
Best Local Similarity 87.0%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgaggg 23  
|||||  
Db 433 GAAATCCATGTAAGGAGGAGGG 455

RESULT 11  
BE019857/c 481 bp mRNA EST 06-JUN-2000  
LOCUS bb60g05.y1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3029912 5'  
DEFINITION similar to TR:O73698 O73698 HYPOTHETICAL 21.5 KD PROTEIN ; , mRNA sequence.  
ACCESSION BE019857  
VERSION BE019857.1 GI:8279946  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 481)  
AUTHORS NTH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC) JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/iresources.shtml

Trace considered overall poor quality

Seq primer: -40RP from Gibco

High quality sequence stop: 1.

#### FEATURES

Location/Qualifiers

1..481

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3029912"

/clone.lib="NIH\_MGC\_9"

/tissue\_type="adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:

ECORI; cDNA made by oligo-dr priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCAGGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 117 a 127 c 138 g 99 t

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 481;

Best Local Similarity 87.0%; Pred. No. 4.8e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaaatccatgtataagcagcaggg 23

||||| ||||| ||||| |||||

Db 259 GAAGTCCATGTAGAGCAACAGGG 237

#### RESULT 12

Az484979/c

LOCUS

DEFINITION 1M0311P16R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0311P16 R. DNA sequence.

ACCESSION Az484979

VERSION Az484979.1 GI:10650348

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

and Wright,D. Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0311 row: P column: 16

Seq primer: CACACGGAACACGATGACC

Class: plasmid ends

High quality sequence stop: 533.

Location/Qualifiers

1..533

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0311P16"

#### FEATURES

source

Location/Qualifiers

1..750

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0311P16"

/clone.lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 139 a 119 c 103 g 172 t

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 13; Length 533;

Best Local Similarity 87.0%; Pred. No. 4.8e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaaatccatgtataagcagcaggg 23

||||| ||||| ||||| |||||

Db 146 GAAGTCTATGTAAGCAGCAGAG 124

#### RESULT 13

Az750976/c

LOCUS

DEFINITION Az750976/c

Mouse 750 bp DNA

musculus genomic clone RPCI-24-128J23

ACCESSION Az750976

VERSION Az750976.1

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Zhao,S., Niemman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,

Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end

plate: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 128 row: J column: 23

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..750

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:55:04 ; Search time 2553.1 Seconds  
(without alignments)  
135.694 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21  
Sequence: 1 gcttgactataacttgac 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hgt: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pi: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pi: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rod: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	6	AX127759
2	19.4	92.4	21	6	AX127451
3	19.4	92.4	249	12	ARGWTUB
c 4	19.4	92.4	415	6	AX127755
5	19.4	92.4	416	6	AX127757
c 6	19.4	92.4	694	6	AX172478
7	19.4	92.4	831	1	ATTRN7
8	19.4	92.4	878	1	ATTDNA
c 9	19.4	92.4	909	6	AX172475
10	19.4	92.4	1037	6	AI0942
c 11	19.4	92.4	1077	6	AX172463
12	19.4	92.4	1085	6	AI0939
13	19.4	92.4	1160	6	AI0943
14	19.4	92.4	1186	6	AI0951
15	19.4	92.4	1186	6	AR095107
16	19.4	92.4	1186	6	AR098313
17	19.4	92.4	1186	6	AX012338
18	19.4	92.4	1186	6	I49886
19	19.4	92.4	1186	6	I82374
c 20	19.4	92.4	2476	12	TBI251013
21	19.4	92.4	3200	6	I44104
22	19.4	92.4	3201	6	I44103
c 23	19.4	92.4	3236	12	TBI251014
c 24	19.4	92.4	4832	6	AX172441
c 25	19.4	92.4	4946	6	A60108
c 26	19.4	92.4	4946	6	A76915
c 27	19.4	92.4	4946	6	AR098307
c 28	19.4	92.4	4946	6	AX172440
c 29	19.4	92.4	5349	6	A71437
c 30	19.4	92.4	5560	6	A60112
c 31	19.4	92.4	5560	6	AR098311
c 32	19.4	92.4	5865	6	AX127748
c 33	19.4	92.4	5865	6	AX127748
34	19.4	92.4	6539	6	E31991
35	19.4	92.4	6548	6	A76916
36	19.4	92.4	6548	6	AR098308
37	19.4	92.4	6548	6	AR098308
38	19.4	92.4	7566	6	A24783
39	19.4	92.4	7566	6	AR074388
40	19.4	92.4	7599	6	AX063413
41	19.4	92.4	7599	6	AX063413
42	19.4	92.4	7639	6	A24782
43	19.4	92.4	7639	6	AR074387
c 44	19.4	92.4	7811	6	AR078675
45	19.4	92.4	12095	12	BINHYGNA

ALIGNMENTS

RESULT 1

AX127759

LOCUS AX127759 21 bp DNA

DEFINITION Sequence 12 from Patent WO0131042.

ACCESSION AX127759

VERSION AX127759.1 GI:14134406

KEYWORDS .

SOURCE synthetic construct.

ORGANISM synthetic construct.

REFERENCE 1 (bases 1 to 21)

AUTHORS Weston,B. and de Beuckeleer,M.

TITLE Male-sterile brassica plants and methods for producing same

JOURNAL Patent: WO 0131042-A 12 03-MAY-2001;

FEATURES

source

1. .21

Location/Qualifiers

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer MDB201"

PAT 15-MAY-2001

BASE COUNT 6 a 4 c 4 g 7 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcttgactataactgtgac 21  
|||||  
Db 1 GCTTGGACTATAACTGTGAC 21

RESULT 2

AX172451 AX172451 21 bp DNA PAT 03-JUL-2001  
LOCUS Sequence 12 from Patent WO0141558.  
DEFINITION AX172451  
ACCESSION AX172451  
VERSION AX172451.1 GI:14597563  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
synthetic construct  
artificial sequence.

REFERENCE 1 (bases 1 to 21)  
AUTHORS de Both,G. and de Beuckeleer,M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 12 14-JUN-2001;  
Aventis CropScience N.V. (BE)

FEATURES  
source  
1..21  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer 201"

BASE COUNT 6 a 5 c 4 g 6 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 21;  
Best Local Similarity 95.2%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataactgtgac 21  
|||||  
Db 1 GCTTGGACTATAACTGTGAC 21

RESULT 3

ARGMTUB ARGMTUB 249 bp DNA SYN 02-APR-1988  
LOCUS Soybean beta-1-tubulin gene fused to T1 plasmid unit 7 3UTR.  
DEFINITION X05579  
ACCESSION X05579  
VERSION X05579.1 GI:58087  
KEYWORDS beta-tubulin; fusion gene; plasmid.  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 249)  
AUTHORS Guiltinan,M.J., Velten,J., Bustos,M.M., Cyr,R.J., Schell,J. and Fosket,D.E.  
TITLE The expression of a chimeric soybean beta-tubulin gene in tobacco  
JOURNAL Mol. Gen. Genet. 207, 328-334 (1987)  
FEATURES  
source  
1..249  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

CDS  
1..51  
/note="fusion product (17AA); Protein sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAA29084.1"  
/db\_xref="GI:4376141"  
/translation="MAAAWASSNWSDPPMS"

misc\_feature 1..13  
/note="beta-1-tubulin sequence"  
misc\_feature 14..28  
/note="pUC 13 polylinker"  
misc\_feature 29..35  
/note="Sal I linker"  
misc\_feature 37  
/note="theroretical fusion junction (24) with gene 7 of T1 plasmid"  
misc\_feature 173..178  
/note="put.polyA signal"  
polyA\_site 198  
/note="polyA site"  
misc\_feature 220..225  
/note="put.polyA signal"  
BASE COUNT 76 a 47 c 38 g 88 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 12; Length 249;  
Best Local Similarity 95.2%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataactgtgac 21  
|||||  
Db 189 GCTTGGACTATAACTGTGAC 209

RESULT 4

AX127755/c AX127755 415 bp DNA PAT 15-MAY-2001  
LOCUS Sequence 8 from Patent WO0131042.  
DEFINITION AX127755  
ACCESSION AX127755  
VERSION AX127755.1 GI:14134402  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
synthetic construct.  
artificial sequence.

REFERENCE 1 (bases 1 to 415)  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 8 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
1..415  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="5' border flanking region of elite event MS-B2"

misc\_feature 1..234  
/note="plant DNA"  
misc\_feature 235..415  
/note="T-DNA"  
BASE COUNT 154 a 55 c 70 g 136 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 415;  
Best Local Similarity 95.2%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataactgtgac 21  
|||||  
Db 335 GCTTGGACTATAACTGTGAC 315

RESULT 5

AX127757 AX127757 416 bp DNA PAT 15-MAY-2001  
LOCUS Sequence 10 from Patent WO0131042.  
DEFINITION AX127757  
ACCESSION AX127757  
VERSION AX127757.1 GI:14134404  
KEYWORDS  
SOURCE  
synthetic construct.

ORGANISM synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 416)  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
1. .416  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="3' border flanking region of elite event MS-B2"  
misc\_feature 1. .193  
/note="T-DNA"  
misc\_feature 194. .416  
/note="plant DNA"  
BASE COUNT 137 a 72 c 54 g 152 t 1 others  
ORIGIN  
  
Query Match 92.4%; Score 19.4; DB 6; Length 416;  
Best Local Similarity 95.2%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 gcttggaactataactgtgac 21  
|||||  
Db 81 GCTTGGACTATAATACCTGAC 101  
  
RESULT 6  
LOCUS AX172478 694 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 39 from Patent WO0141558.  
ACCESSION AX172478  
VERSION AX172478.1 GI:14597590  
KEYWORDS  
SOURCE  
ORGANISM synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 694)  
AUTHORS de Both,G. and de Beuckeleer,M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 39 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
1. .694  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="sequence comprising the 5' flanking region of RF-BN1 in WOSR"  
BASE COUNT 298 a 81 c 73 g 242 t  
ORIGIN  
  
Query Match 92.4%; Score 19.4; DB 6; Length 694;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 gcttggaactataactgtgac 21  
|||||  
Db 694 GCTTGGACTATAATACCTGAC 674  
  
RESULT 7  
LOCUS ATTRN7 831 bp DNA BCT 02-SEP-1999  
DEFINITION Agrobacterium tumefaciens gene encoding 'transcript 7' mRNA for a protein with unknown function.  
ACCESSION V00090  
VERSION V00090.1 GI:39180  
KEYWORDS unidentified reading frame.  
SOURCE Agrobacterium tumefaciens.  
ORGANISM Agrobacterium tumefaciens

Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Rhizobiaceae; Rhizobium.  
REFERENCE 1 (bases 1 to 831)  
AUTHORS Dhaese,P., De Greve,H., Gielen,J., Seurinck,J., Van Montagu,M.M.  
and Schell,J.  
TITLE Identification of sequences involved in the polyadenylation of  
higher plant nuclear transcripts using Agrobacterium T-DNA genes as  
models  
JOURNAL EMBO J. 2, 419-426 (1983)  
REFERENCE 2 (bases 76 to 100)  
AUTHORS Dhaese,P.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAY-1983) to the EMBL/GenBank/DBJ databases  
COMMENT Data kindly reviewed (27-MAY-1983) by Dhaese P.  
FEATURES  
Location/Qualifiers  
source  
1. .831  
/organism="Agrobacterium tumefaciens"  
/strain="(octopine TL-DNA)"  
/db\_xref="taxon:358"  
75. .81  
/note="CG [1] revised CCAGAGG [2]"  
/citation=[1]  
/citation=[2]  
99. .101  
/note="CCA [1] revised CTA [2]"  
/citation=[1]  
/citation=[2]  
132. .673  
/note="transcript 7 (alternate)"  
132. .672  
/note="transcript 7"  
148. .528  
/note="unknown gene (148 is 1st base in codon) (525 is 3rd  
base in codon)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA23429.1"  
/db\_xref="GI:39181"  
/db\_xref="SWISS-PROT:P03867"  
/translation="MNEADTPLASLDLDWACEFFIKTYGASPOLETGEVIOIOTNNGLLY  
LYKGSLSQRIHDTLHFKFEKEELSFTTKPAEMKAQOSDLYTVVAIFQSNFLCVSN  
PEKFLRCHNRPFPIPIVAHGSMS"  
BASE COUNT 262 a 176 c 138 g 255 t  
ORIGIN  
  
Query Match 92.4%; Score 19.4; DB 1; Length 831;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 gcttggaactataactgtgac 21  
|||||  
Db 664 GCTTGGACTATAATACCTGAC 684  
  
RESULT 8  
LOCUS ATTDNA 878 bp DNA BCT 25-MAR-1996  
DEFINITION Agrobacterium tumefaciens crown gall tumor T-DNA from T1 (tumor  
inducing) plasmid pTiA6.  
ACCESSION X00431  
VERSION X00431.1 GI:39150  
KEYWORDS plasmid.  
SOURCE Agrobacterium tumefaciens.  
ORGANISM Agrobacterium tumefaciens  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Rhizobiaceae; Rhizobium.  
REFERENCE 1 (bases 1 to 878)  
AUTHORS McPherson,J.C.  
TITLE DNA sequence analysis of crown gall tumor T-DNA encoding the 0.7 kb  
transcript  
JOURNAL Nucleic Acids Res. 12 (5), 2317-2325 (1984)  
MEDLINE 84169535

```

FEATURES
  source          Location/Qualifiers
    1..878
      /organism="Agrobacterium tumefaciens"
      /strain="plasmid pTIA6"
      /db_xref="taxon:358"
    60..66
      /note="TATA-box"
    68..613
      /note="polyadenylation signal"
    109..489
      /note="unidentified reading frame"
      /codon_start=1
      /transl_table=11
      /protein_id="CAA25129.1"
      /db_xref="GI:39151"
      /db_xref="SWISS-PROT:P03867"
      /translation="MNFADTPLASLDLDMACEEFIKTYGASPOLETGEVIQTNNGLLY
      LYKGSLSQRHDTHLKPKKEELSFTTIKPAEMKAQSDLYVVAIFQSNFLCVSN
      PKGFLRCHNRPFLLPIVAHGSM"
    misc_feature   659..664
      /note="polyadenylation signal"
  BASE COUNT      288 a 189 c 139 g 262 t
  ORIGIN
    1 gcttgactataacttgac 21
    |||||
  Db 625 GCTTGGACTATAATACCTGAC 645

  Query Match      92.4%; Score 19.4; DB 1; Length 878;
  Best Local Similarity 95.2%; Pred. No. 11;
  Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 1 gcttgactataacttgac 21
  |||||
  Db 625 GCTTGGACTATAATACCTGAC 645

RESULT 9
AXI172475/c
LOCUS AXI172475 909 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 36 from Patent WO0141558.
ACCESSION AXI172475
VERSION AXI172475.1 GI:14597587
KEYWORDS
SOURCE
  ORGANISM
    synthetic construct.
    artificial construct.
  REFERENCE
    1 (bases 1 to 909)
    de Both.G. and de Beuckeleer.M.
    Hybrid winter oilseed rape and methods for producing same
    Patent: WO 0141558-A 36 14-JUN-2001;
    Aventis CropScience N.V. (BE)
  FEATURES
    source
      Location/Qualifiers
        1..909
          /organism="synthetic construct"
          /db_xref="taxon:32630"
          /note="sequence comprising the 5' flanking region of
          MS-BN1 in WOSR"
  BASE COUNT      309 a 134 c 173 g 293 t
  ORIGIN
    1 gcttgactataacttgac 21
    |||||
  Db 909 GCTTGGACTATAATACCTGAC 889

  Query Match      92.4%; Score 19.4; DB 6; Length 909;
  Best Local Similarity 95.2%; Pred. No. 11;
  Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 1 gcttgactataacttgac 21
  |||||
  Db 909 GCTTGGACTATAATACCTGAC 889

RESULT 10
AXI0942
LOCUS AXI0942 1037 bp DNA PAT 27-SEP-1993
DEFINITION Nucleotide sequence 4 from patent number DE3920034.
ACCESSION AXI0942

```

```

  VERSION          A10942.1 GI:492369
  KEYWORDS          .
  SOURCE            unidentified.
  ORGANISM          unidentified.
  REFERENCE        1 (bases 1 to 1037)
  JOURNAL          Patent: DE 3920034-A 4 31-MAY-1990;
  FEATURES          Location/Qualifiers
    source          1..1037
      /organism="unidentified"
      /db_xref="taxon:32644"
  BASE COUNT      338 a 174 c 166 g 359 t
  ORIGIN
    1 gcttgactataacttgac 21
    |||||
  Db 921 GCTTGGACTATAATACCTGAC 941

  Query Match      92.4%; Score 19.4; DB 6; Length 1037;
  Best Local Similarity 95.2%; Pred. No. 11;
  Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 1 gcttgactataacttgac 21
  |||||
  Db 921 GCTTGGACTATAATACCTGAC 941

RESULT 11
AXI172463/c
LOCUS AXI172463 1077 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 24 from Patent WO0141558.
ACCESSION AXI172463
VERSION AXI172463.1 GI:14597575
KEYWORDS
SOURCE
  ORGANISM
    synthetic construct.
    artificial construct.
  REFERENCE
    1 (bases 1 to 1077)
    de Both.G. and de Beuckeleer.M.
    Hybrid winter oilseed rape and methods for producing same
    Patent: WO 0141558-A 24 14-JUN-2001;
    Aventis CropScience N.V. (BE)
  FEATURES
    source
      Location/Qualifiers
        1..1077
          /organism="synthetic construct"
          /db_xref="taxon:32630"
          /note="sequence comprising the 5' flanking region of
          RF-BN1"
  misc_feature     1..45
      /note="pGEM -T vector"
  misc_feature     1061..1077
      /note="pGEM -T vector"
  BASE COUNT      430 a 148 c 151 g 347 t 1 others
  ORIGIN
    1 gcttgactataacttgac 21
    |||||
  Db 980 GCTTGGACTATAATACCTGAC 960

  Query Match      92.4%; Score 19.4; DB 6; Length 1077;
  Best Local Similarity 95.2%; Pred. No. 11;
  Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 1 gcttgactataacttgac 21
  |||||
  Db 980 GCTTGGACTATAATACCTGAC 960

RESULT 12
AXI0939
LOCUS AXI0939 1085 bp DNA PAT 27-SEP-1993
DEFINITION Nucleotide sequence 1 from patent number DE3920034.
ACCESSION AXI0939
VERSION AXI0939.1 GI:492367
KEYWORDS
SOURCE
  ORGANISM
    unidentified.
    unclassified.

```



Search completed: February 15, 2002, 18:55:04  
Job time: 20543 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:02:02 ; Search time 868.33 Seconds  
(without alignments)  
20.734 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21

Sequence: 1 gcttgactataacttgac 21

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N\_Geneseq\_1101:\*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:\*
  - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:\*
  - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:\*
  - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:\*
  - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:\*
  - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:\*
  - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:\*
  - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:\*
  - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:\*
  - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:\*
  - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:\*
  - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:\*
  - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:\*
  - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:\*
  - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:\*
  - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:\*
  - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:\*
  - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:\*
  - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:\*
  - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:\*
  - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:\*
  - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	21	100.0	21	22	AA07001
2	19.4	92.4	21	22	AAH25431
3	19.4	92.4	415	22	AA006997
4	19.4	92.4	416	22	AA006999
5	19.4	92.4	694	22	AAH25454
6	19.4	92.4	909	22	AAH25451
7	19.4	92.4	1037	11	AA004705
8	19.4	92.4	1077	22	AAH25439
9	19.4	92.4	1085	11	AA004703
10	19.4	92.4	1160	11	AA004706
11	19.4	92.4	1186	13	AAQ25707

12	19.4	92.4	1303	17	AAAT39337
13	19.4	92.4	3153	21	AAZ29122
14	19.4	92.4	3201	12	AAQ14529
15	19.4	92.4	3201	12	AAQ15144
16	19.4	92.4	3336	21	AAZ29121
17	19.4	92.4	3694	21	AAZ29124
18	19.4	92.4	3877	21	AAZ29123
19	19.4	92.4	4832	22	AAH25423
20	19.4	92.4	4946	18	AAAT59531
21	19.4	92.4	4946	22	AAH25422
22	19.4	92.4	5228	22	AAAF86439
23	19.4	92.4	5349	19	AAZ23239
24	19.4	92.4	5864	17	AAAT39339
25	19.4	92.4	5864	17	AAAT39339
26	19.4	92.4	5865	22	AAAD06990
27	19.4	92.4	5865	22	AAAD06990
28	19.4	92.4	6539	21	AAZ91097
29	19.4	92.4	6548	17	AAAT39336
30	19.4	92.4	6548	18	AAAT61394
31	19.4	92.4	6548	21	AAZ91096
32	19.4	92.4	7492	22	AAAF86441
33	19.4	92.4	7566	14	AAQ42160
34	19.4	92.4	7599	22	AAAF25320
35	19.4	92.4	7639	14	AAQ42159
36	19.4	92.4	24593	6	AAAN50226
37	19.4	92.4	24596	6	AAAN50182
38	16.2	77.1	423	21	AAAC02009
39	16	75.2	22	20	AAZ28158
40	15.8	75.2	81	16	AAAT06276
41	15.8	75.2	359	21	AAAC06027
42	15.8	75.2	831	20	AAZ10777
43	15.8	75.2	909	20	AAZ22113
44	15.8	75.2	1433	15	AAO65397
45	15.8	75.2	1433	16	AAQ92306

ALIGNMENTS

RESULT 1

AA07001  
ID AAD07001 standard; DNA; 21 BP.  
AC AAD07001;  
XX  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE PCR primer B02, to recognise foreign DNA and flanking sequence of MS-B2.  
XX  
KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; PCR primer; ss.  
XX  
OS Unidentified.  
XX  
FN WO200131042-A2.  
XX  
PD 03-MAY-2001.  
XX  
PF 26-OCT-2000; 2000WO-EP10680.  
XX  
PR 29-OCT-1999; 99US-0430497.  
XX  
XX (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
XX Weston B, De Beuckeleer M;  
XX  
DR WPI; 2001-300517/31.  
XX  
PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
XX male-sterility gene, at specific location in its genome -  
PS Claim 1; Page 33; 53pp; English.

```

XX CC The present invention relates to a transgenic Brassica plant or its
CC seed, cells or tissues, characterised by harbouring a specific
CC transformation event, particularly by the presence of a male-sterility
CC gene, at a specific location in the Brassica genome. Transgenic
CC Brassica plant is useful for producing a hybrid seed by crossing the
CC transgenic plant with a male-fertile Brassica plant and harvesting the
CC hybrid seed from the transgenic Brassica plant.
CC The present sequence is PCR primer which is used to recognise foreign
CC DNA and a flanking sequence of elite event MS-B2.
XX SQ Sequence 21 BP; 6 A; 4 C; 4 G; 7 T; 0 other;

Query Match      100.0%; Score 21; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21
Db 1 gcttggactataacttgac 21
      |||||
RESULT 2
AAH25431
ID AAH25431 standard; DNA; 21 BP.
XX AC AAH25431;
XX FH Key Location/Qualifiers
XX FT misc_feature 1..234
XX FT /*tag= a
XX FT /note= "Corresponds to plant DNA"
XX FT misc_feature 235..415
XX FT /*tag= b
XX FT /note= "Corresponds to T-DNA"
XX PN WO200131042-A2.
XX PD 03-MAY-2001.
XX XX
XX PF 26-OCT-2000; 2000WO-EP10680.
XX XX
XX PR 29-OCT-1999; 99US-0430497.
XX XX
XX PA (AVET ) AVENTIS CROPS SCIENCE NV.
XX XX
XX PI De Both G, De Beuckeleer M;
XX DR WPI; 2001-381419/40.
XX XX
XX PT Transgenic winter oilseed rape plants suited for producing hybrid seed
XX PT with improved qualities, comprises a male-sterility gene and fertility
XX PT restorer gene, integrated into the genome -
XX PS Example 4; Page 43; 98pp; English.
XX CC The specification describes a pair of transgenic winter oilseed rape
XX CC plants suited for producing hybrid seed. One of the plants has an
XX CC expression cassette comprising a male-sterility gene (e.g. barnase
XX CC gene), and the other plant has an expression cassette comprising a
XX CC fertility restorer gene (e.g. barstar gene), integrated into the genome.
XX CC The fertility restorer gene is capable of preventing the activity of the
XX CC male-sterility gene. The plant pair is useful for producing hybrid seed.
XX CC Plants developed from the hybrid seed have agronomic performance,
XX CC genetic stability and adaptability to different genetic backgrounds.
XX CC PCR primers AAH25429-31 were used to amplify fragments of a vector in
XX CC a transgenic plant which carries the TA29-barnase transgene.
XX SQ Sequence 21 BP; 6 A; 5 C; 4 G; 6 T; 0 other;

```

```

Query Match      92.4%; Score 19.4; DB 22; Length 21;
Best Local Similarity 95.2%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21
Db 1 gcttggactataacttgac 21
      |||||
RESULT 3
AAD06997/c
ID AAD06997 standard; DNA; 415 BP.
XX AC AAD06997;
XX DT 06-AUG-2001 (first entry)
XX DE Right (5') border flanking region of elite event MS-B2.
XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;
XX KW male-sterility gene; ds.
XX OS Chimeric - Agrobacterium sp.
XX OS Chimeric - Brassica sp.
XX FH Key Location/Qualifiers
XX FT misc_feature 1..234
XX FT /*tag= a
XX FT /note= "Corresponds to plant DNA"
XX FT misc_feature 235..415
XX FT /*tag= b
XX FT /note= "Corresponds to T-DNA"
XX PN WO200131042-A2.
XX PD 03-MAY-2001.
XX XX
XX PF 26-OCT-2000; 2000WO-EP10680.
XX XX
XX PR 29-OCT-1999; 99US-0430497.
XX XX
XX PA (AVET ) AVENTIS CROPS SCIENCE NV.
XX XX
XX PI Weston B, De Beuckeleer M;
XX DR WPI; 2001-300517/31.
XX XX
XX PT Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX PT harbouring specific transformation events, particularly by presence of
XX PT male-sterility gene, at specific location in its genome -
XX PS Claim 11; Page 51; 53pp; English.
XX CC The present invention relates to a transgenic Brassica plant or its
XX CC seed, cells or tissues, characterised by harbouring a specific
XX CC transformation event, particularly by the presence of a male-sterility
XX CC gene, at a specific location in the Brassica genome. Transgenic
XX CC Brassica plant is useful for producing a hybrid seed by crossing the
XX CC transgenic plant with a male-fertile Brassica plant and harvesting the
XX CC hybrid seed from the transgenic Brassica plant.
XX CC The present sequence is right (5') border flanking region of elite event
XX CC MS-B2.
XX SQ Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match      92.4%; Score 19.4; DB 22; Length 415;
Best Local Similarity 95.2%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21
Db 335 GCTTGGACTATAATACCTGAC 315
      |||||

```

```

RESULT 4
AAD06999
ID AAD06999 standard; DNA; 416 BP.
XX
AC AAD06999;
XX
DT 06-AUG-2001 (first entry)
XX
DE Left (3') border flanking region of elite event MS-B2.
XX
KW MS-B2 elite event; transgenic Brassica plant; transformation event;
KW male-sterility gene; ds.
XX
OS Chimeric - Agrobacterium sp.
OS Chimeric - Brassica sp.
XX
FH Key Location/Qualifiers
FT misc_feature 1..193
FT /*tag= a
FT /*note= "Corresponds to T-DNA"
FT misc_feature 194..416
FT /*tag= b
FT /*note= "Corresponds to plant DNA"
XX
XX WO200131042-A2.
XX
XX 03-MAY-2001.
XX
XX 26-OCT-2000; 2000WO-EP10680.
XX
XX 29-OCT-1999; 99US-0430497.
XX
XX (AVET ) AVENTIS CROPSCIENCE NV.
XX
XX Weston B, De Beuckeleer M;
XX
XX WPI; 2001-300517/31.
XX
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX harboring specific transformation events, particularly by presence of
XX male-sterility gene, at specific location in its genome -
XX
XX Claim 11; Page 52; 53pp; English.
XX
XX The present invention relates to a transgenic Brassica plant or its
XX seed, cells or tissues, characterised by harbouring a specific
XX transformation event, particularly by the presence of a male-sterility
XX gene, at a specific location in the Brassica genome. Transgenic
XX Brassica plant is useful for producing a hybrid seed by crossing the
XX transgenic plant with a male-fertile Brassica plant and harvesting the
XX hybrid seed from the transgenic Brassica plant.
XX The present sequence is left (3') border flanking region of elite event
XX MS-B2.
XX
XX Sequence 416 BP; 137 A; 72 C; 54 G; 152 T; 1 other;

Query Match 92.4%; Score 19.4; DB 22; Length 416;
Best Local Similarity 95.2%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21
| | | | | | | | | | | | | | | |
Db 81 gcttggactataacttgac 101

RESULT 5
AAH25454/C
ID AAH25454 standard; DNA; 694 BP.
XX
AC AAH25454;

Query Match 92.4%; Score 19.4; DB 22; Length 416;
Best Local Similarity 95.2%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21
| | | | | | | | | | | | | | | |
Db 81 gcttggactataacttgac 101

```

```

XX 22-AUG-2001 (first entry)
XX Nucleotide fragment of a vector comprising TA29-barstar transgene.
DE Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
XX fertility restorer gene; barnase gene; barstar gene; ss.
XX Synthetic.
XX WO200141558-A1.
XX 14-JUN-2001.
XX 06-DEC-2000; 2000WO-EP12872.
XX 08-DEC-1999; 99US-0457037.
XX (AVET ) AVENTIS CROPSCIENCE NV.
XX De Both G, De Beuckeleer M;
XX WPI; 2001-381419/40.
XX Transgenic winter oilseed rape plants suited for producing hybrid seed
XX with improved qualities, comprises a male-sterility gene and fertility
XX restorer gene, integrated into the genome -
XX Example 4; Page 94; 98pp; English.
XX The specification describes a pair of transgenic winter oilseed rape
XX plants suited for producing hybrid seed. One of the plants has an
XX expression cassette comprising a male-sterility gene (e.g. barnase
XX gene) and the other plant has an expression cassette comprising a
XX fertility restorer gene (e.g. barstar gene), integrated into the genome.
XX The fertility restorer gene is capable of preventing the activity of the
XX male-sterility gene. The plant pair is useful for producing hybrid seed.
XX Plants developed from the hybrid seed have agronomic performance,
XX genetic stability and adaptability to different genetic backgrounds.
XX The present sequence represents a fragment from a vector comprising the
XX TA29-barstar transgene from transgenic plants.
XX Sequence 694 BP; 298 A; 81 C; 73 G; 242 T; 0 other;

Query Match 92.4%; Score 19.4; DB 22; Length 694;
Best Local Similarity 95.2%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21
| | | | | | | | | | | | | | | |
Db 694 GCTTGGACTATAACTTGAC 674

RESULT 6
AAH25451/C
ID AAH25451 standard; DNA; 909 BP.
XX
AC AAH25451;
XX
XX 22-AUG-2001 (first entry)
XX
XX Nucleotide fragment of a vector comprising TA29-barnase transgene.
DE Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barnase gene; barstar gene; ss.
XX Synthetic.
XX WO200141558-A1.
XX 14-JUN-2001.
XX

```

PF 06-DEC-2000; 2000WO-EP12872.  
 PR 08-DEC-1999; 99US-0457037.  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA De Both G, De Beuckeleer M;  
 XX WPI; 2001-381419/40.  
 DR  
 XX Transgenic winter oilseed rape plants suited for producing hybrid seed  
 PT with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome -  
 XX Example 4; Page 93; 98pp; English.  
 XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC The present sequence represents a fragment from a vector comprising the  
 CC TA29-barnase transgene from transgenic plants.  
 XX Sequence 909 BP; 309 A; 134 C; 173 G; 293 T; 0 other;

Query Match 92.4%; Score 19.4; DB 22; Length 909;  
 Best Local Similarity 95.2%; Pred. No. 1.3;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcttgactataactgtgac 21  
 |||||  
 DB 909 GCTTGACTATAACTGTGAC 889

RESULT 7  
 AAQ04705  
 ID AAQ04705 standard; DNA; 1037 BP.  
 XX  
 AC AAQ04705;  
 XX  
 DT 12-OCT-1990 (first entry)  
 XX  
 XX USP-Promoter-cassette USP-Pr.T7.1.  
 XX Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.  
 KW  
 XX DE3920034-A.  
 XX  
 XX 31-MAY-1990.  
 XX  
 XX 20-JUN-1989; 89DE-3920034.  
 XX  
 XX 19-SEP-1988; 88DD-0319887.  
 XX  
 XX (PFLA-) VE KOMB PFLANZENZUC.  
 XX  
 XX Basseuner R, Baumllein H, Muntz K, Hai NV, Wobus U;  
 PI WPI; 1990-172459/23.  
 XX  
 XX  
 XX  
 PT Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.  
 XX  
 XX Disclosure; ; pp; German.  
 XX  
 XX The unique BglII-Ort (720-725) site is for ligating foreign DNA and the

CC HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 CC cassette in the Ti-vector pGA471. Agrobacterium tumefaciens is  
 CC transfected.  
 CC See also AAQ04703-Q04706.  
 XX  
 SQ Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 92.4%; Score 19.4; DB 11; Length 1037;  
 Best Local Similarity 95.2%; Pred. No. 1.4;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataactgtgac 21  
 |||||  
 DB 921 gcttgactataactgtgac 941

RESULT 8  
 AAH25439/C  
 ID AAH25439 standard; DNA; 1077 BP.  
 XX  
 AC AAH25439;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 XX Right flanking region in transgenic plant BN-RF1.  
 XX  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; barstar gene; ss.  
 XX  
 OS Synthetic.  
 XX  
 XX WO200141558-A1.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 06-DEC-2000; 2000WO-EP12872.  
 XX  
 PR 08-DEC-1999; 99US-0457037.  
 XX  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA  
 XX De Both G, De Beuckeleer M;  
 PI  
 XX WPI; 2001-381419/40.  
 XX  
 XX Transgenic winter oilseed rape plants suited for producing hybrid seed  
 PT with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome -  
 XX Claim 58; Page 88-89; 98pp; English.  
 XX  
 XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC The present sequence represents the right flanking region of  
 CC a vector in a transgenic plant which carries the TA29-barstar  
 CC transgene.  
 XX  
 SQ Sequence 1077 BP; 430 A; 148 C; 151 G; 347 T; 1 other;

Query Match 92.4%; Score 19.4; DB 22; Length 1077;  
 Best Local Similarity 95.2%; Pred. No. 1.4;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcttgactataactgtgac 21

```

Db 980 GCTTGGACTATAATACCTGC 960
      |||||
RESULT 9
AAQ04703
ID AAQ04703 standard; DNA; 1085 BP.
XX AC AAQ04703;
XX DT 12-OCT-1990 (first entry)
XX DE Legumin-signalpeptide cassette Le-Sig.T7.
XX KW Foreign DNA incorporation; recombinant DNA techniques;
XX KW higher plant genome; legumin; B4 gene; signalpeptide; Le-Sig.T7.; ss.
XX XX
FH Key Location/Qualifiers
FT CDS 747..814
FT /*tag= a
FT /product=Legumin-signalpeptide
XX DE3920034-A.
XX PN 31-MAY-1990.
XX PF 20-JUN-1989; 89DE-3920034.
XX PR 19-SEP-1988; 88DD-0319887.
XX PA (PFLA-) VE KOMB PFLAN2ENZUC.
XX PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;
XX DR WPI; 1990-172459/23.
XX DR P-PSDB; AAR05198.
XX PT Incorporation of DNA into higher plant genome - by specified
XX PT recombinant DNA techniques.
XX PS Disclosure; ; pp; German.
XX CC The Legumin gene B4 is used. The unique BglII-Ort (815-820) site is for
XX CC ligating foreign DNA and the HindIII-Ort in the 3' polylinker
XX CC (1080-1085) for cloning the cassette in the Ti-vector PGA471.
XX CC The cassette is cloned into the binary Ti-vectors PGA471 and
XX CC Agrobacterium tumefaciens is transfected.
XX CC See also AAQ04703-Q04706.
XX SQ Sequence 1085 BP; 368 A; 219 C; 155 G; 343 T; 0 other;

Query Match 92.4%; Score 19.4; DB 11; Length 1085;
Best Local Similarity 95.2%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttggactataactctgac 21
   |||||
Db 969 gcttggactataactctgac 989

RESULT 10
AAQ04706
ID AAQ04706 standard; DNA; 1160 BP.
XX AC AAQ04706;
XX DT 12-OCT-1990 (first entry)
XX DE USP-signalpeptide cassette USP-Sig.T7.
XX KW Foreign DNA incorporation; recombinant DNA techniques;
XX KW higher plant genome; signalpeptide; USP-Sig.T7.; ss.

```

```

XX FH Key Location/Qualifiers
XX FT CDS 708..877
XX FT /*tag= a
XX FT /product=signalpeptide
XX FT 747..817
XX PN DE3920034-A.
XX PD 31-MAY-1990.
XX PF 20-JUN-1989; 89DE-3920034.
XX PR 19-SEP-1988; 88DD-0319887.
XX PA (PFLA-) VE KOMB PFLAN2ENZUC.
XX PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;
XX DR WPI; 1990-172459/23.
XX DR P-PSDB; AAR05199.
XX PT Incorporation of DNA into higher plant genome - by specified
XX PT recombinant DNA techniques.
XX PS Disclosure; ; pp; German.
XX CC The unique BglII-Ort (890-895) site is for
XX CC ligating foreign DNA and the HindIII-Ort in the 3' polylinker
XX CC (1155-1160) for cloning the cassette in the Ti-vector PGA471.
XX CC The cassette is cloned into the binary Ti-vectors PGA471 and
XX CC Agrobacterium tumefaciens is transfected.
XX CC See also AAQ04703-Q04706.
XX SQ Sequence 1160 BP; 369 A; 192 C; 188 G; 411 T; 0 other;

Query Match 92.4%; Score 19.4; DB 11; Length 1160;
Best Local Similarity 95.2%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttggactataactctgac 21
   |||||
Db 1044 gcttggactataactctgac 1064

RESULT 11
AAQ25707
ID AAQ25707 standard; DNA; 1186 BP.
XX AC AAQ25707;
XX DT 07-DEC-1992 (first entry)
XX DE Chimeric neo gene probe.
XX KW Nicotiana tabacum; neomycine phosphotransferase; Agrobacterium; ss.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT promoter 1..8
XX FT /*tag= a
XX FT /note= "sequence derived from tapetum specific
XX FT promoter of Nicotiana tabacum"
XX FT CDS 9..790
XX FT /*tag= b
XX FT /product= neomycine_phosphotransferase
XX FT misc_feature 791..1186
XX FT /*tag= c
XX FT /note= "3' regulatory sequence contg. the
XX FT polyadenylation site derived from
XX FT Agrobacterium T-DNA gene 7"

```

```

XX PN WO9209696-A.
XX PD 11-JUN-1992.
XX PF 21-NOV-1991; 91WO-EP02198.
XX PR 23-NOV-1990; 90EP-0403332.
XX PR 08-JUL-1991; 91EP-0401888.
XX PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PI Dhalluin K, Goebel E;
XX PD WPI; 1992-217075/26.
XX DR Transforming monocotyledonous plants e.g. cereals - comprises
XX PT wounding and/or degrading cells of intact plant tissue or
XX PT embryogenic callus
XX PS Disclosure; Page 60; 76pp; English.
XX CC Two transformed corn plants were analysed by means of Southern
XX CC hybridisation. As a probe, a 1184 bp EcoRI-HindIII fragment derived
XX CC from another plasmid was used. The sequence of that plasmid is
XX CC given below. Results showed that at least a chimeric neo gene was
XX CC integrated into the plant genomic DNA.
XX CC Sequence 1186 BP; 244 A; 317 C; 325 G; 300 T; 0 other;
XX SQ

Query Match 92.4%; Score 19.4; DB 13; Length 1186;
Best Local Similarity 95.2%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataataactgac 21
   |||||
DB 1113 gcttgactataataactgac 1133

RESULT 12
AAT39337
ID AAT39337 standard; DNA; 1303 BP.
XX AC AAT39337;
XX DT 22-JAN-1997 (first entry)
XX DE Plasmid pTS88 (EcoRI-HindIII fragment).
XX KW Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
XX KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT misc_feature 1..35
XX FT /*tag= a
XX FT /label= pGEM2
XX FT /note= "polylinker of pGEM2"
XX FT promoter 36..694
XX FT /*tag= b
XX FT /label= P35S
XX FT /function= 35S promoter of cauliflower mosaic virus
XX CDS 695..967
XX FT /*tag= c
XX FT /label= barstar
XX FT /product= Bacillus amyloliquefaciens barstar
XX FT 968..1287
XX FT /*tag= d
XX FT /label= 3'g7
XX FT /function= region containing polyadenylation signal

```

```

FT misc_feature 1288..1303 of gene 7 og Agrobacterium T-DNA
FT /*tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX PN WO9626283-A1.
XX PD 29-AUG-1996.
XX PF 21-FEB-1996; 96WO-EP00722.
XX PR 21-FEB-1995; 95EP-0400364.
XX PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PI Botterman J, Cornelissen M, Michiels F;
XX PD WPI; 1996-402373/40.
XX CC Prodn. of male sterile plants by transforming with a chimaeric
XX CC construct - comprising a male sterility DNA e.g. barnase and a
XX CC co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX CC for generating hybrid cultivars
XX PS Example 1; Page 38; 56pp; English.
XX CC The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains
XX CC barstar DNA under control of a 35S promoter. The plasmid was
XX CC used with pTS174 (see also AAT39336) contg. barnase DNA under
XX CC control of the stamen-specific promoter E1 to produce male sterile
XX CC rice cv. Kochinibiki transgenic plants, and with plasmid pVE136
XX CC (see also AAT39338) contg. barnase DNA under control of the stamen-
XX CC specific pCA55 promoter to produce male sterile maize plants.
XX CC Expression of barnase (a ribonuclease) in the stamen leads to male
XX CC sterility. Constitutive expression of barstar counteracts possible
XX CC low level expression of barnase DNA in non-stamen tissue.
XX SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 92.4%; Score 19.4; DB 17; Length 1303;
Best Local Similarity 95.2%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataataactgac 21
   |||||
DB 1178 gcttgactataataactgac 1198

RESULT 13
AAZ29122
ID AAZ29122 standard; DNA; 3153 BP.
XX AC AAZ29122;
XX DT 21-FEB-2000 (first entry)
XX DE Plasmid DV131 comprising L3/EGFP:NPTII/Tr7 expression cassette.
XX KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DV131;
XX KW GFP:NPTII fusion protein construct; L3/EGFP:NPTII/Tr7; maize R gene;
XX KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;
XX KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;
XX KW Green fluorescent protein; GFP; gene fusion; selection; screening;
XX KW expression; automated seed screening technique; screenable marker;
XX KW transformant; embryogenic tissue; implementation; ds.
XX OS Synthetic.
XX PN WO9960129-A1.
XX PD 25-NOV-1999.

```



PT	Modified Bacillus thuringiensis insecticidal crystal protein genes - having A and T sequences changed to G and C sequences encoding same amino acids, for increased expression levels
PT	XX
PT	XX
PS	Disclosure; Fig 6(c); 78pp; English.
XX	XX
CC	"n" in the sequence refers to not known nucleotides.
CC	pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-terminal modification and the internal modification of the Bt ICP coding sequence.
CC	See also AAQ14529, AAQ15142-44.
CC	XX
XX	Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;
SQ	
Query Match            92.4%;    Score 19.4;    DB 12;    Length 3201;	
Best Local Similarity    95.2%;    Pred. No. 1.5;	
Matches    20; Conservative    0; Mismatches    1; Indels    0; Gaps	
QY	1 gcttggaactataatcattgac 21 
Db	3032 gcttggaactataatcattgac 3052
RESULT 15	
AAQ15144	
ID	AAQ15144 standard; DNA; 3201 BP.
XX	XX
AC	AAQ15144;
XX	XX
DT	27-JAN-1992 (first entry)
XX	XX
DE	pVE36 Bt ICP coding sequence.
XX	XX
KW	Bacillus thuringiensis; insecticidal crystal protein; ICP; deletion; ss.
KW	XX
XX	Synthetic.
OS	XX
XX	WO9116432-A.
PN	XX
PD	31-OCT-1991.
XX	XX
PF	17-APR-1991;    91WO-EP00733.
PR	18-APR-1990;    90EP-0401055.
XX	(PLAN-) PLANT GENETIC SYST.
PA	XX
XX	Cornelissen M, Soetaert P, Stam M, Dockx J;
PI	WPI; 1991-339820/46.
DR	XX
XX	Modified Bacillus thuringiensis insecticidal crystal protein genes - having A and T sequences changed to G and C sequences encoding same amino acids, for increased expression levels
PT	XX
PT	XX
PS	Disclosure; Fig 6(c); 78pp; English.
XX	XX
CC	"n" in the sequence refers to not known nucleotides.
CC	pPS029 (AAQ14529) is identical to pVE36, but carries both the amino-terminal modification and the internal modification of the Bt ICP coding sequence.
CC	See also AAQ14529, AAQ15142-44.
CC	XX
XX	Sequence 3201 BP; 915 A; 651 C; 700 G; 930 T; 5 other;
SQ	
Query Match            92.4%;    Score 19.4;    DB 12;    Length 3201;	
Best Local Similarity    95.2%;    Pred. No. 1.5;	
Matches    20; Conservative    0; Mismatches    1; Indels    0; Gaps	
QY	1 gcttggaactataatcattgac 21

Tue Feb 19 10:58:32 2002

us-09-698-903b-12.rng

Page 8

Db 3105 gcttgactataacctgac 3125  
|||||

Search completed: February 15, 2002, 19:02:03  
Job time: 20767 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:09:16 ; Search time 353.79 Seconds  
(without alignments)  
13.443 Million cell updates/sec

Title: US-09-698-903b-12

Perfect score: 21  
Sequence: 1 gcttgactataacttgac 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	92.4	1186	1	US-08-064-121-2
2	19.4	92.4	1186	1	US-08-478-015-2
3	19.4	92.4	1186	3	US-08-475-975-2
4	19.4	92.4	1186	3	US-09-084-889-2
5	19.4	92.4	1303	3	US-08-894-440-2
6	19.4	92.4	3153	4	US-09-080-625-3
7	19.4	92.4	3200	1	US-08-453-104-23
8	19.4	92.4	3200	2	US-08-894-824-23
9	19.4	92.4	3201	1	US-08-453-104-22
10	19.4	92.4	3201	2	US-08-694-824-22
11	19.4	92.4	3336	4	US-09-080-625-2
12	19.4	92.4	3694	4	US-09-080-625-5
13	19.4	92.4	3877	4	US-09-080-625-4
14	19.4	92.4	4946	3	US-08-817-188-1
15	19.4	92.4	5560	3	US-08-817-188-5
16	19.4	92.4	5864	3	US-08-894-440-4
17	19.4	92.4	5864	3	US-08-894-440-1
18	19.4	92.4	6548	3	US-08-894-440-1
19	19.4	92.4	6548	3	US-08-817-188-2
20	19.4	92.4	7566	2	US-08-232-016-23
21	19.4	92.4	7639	2	US-08-232-016-22
22	19.4	92.4	7811	2	US-08-549-680A-5
23	19.4	92.4	24595	6	5428147-1
24	15.8	75.2	81	1	US-08-238-863-88
25	15.8	75.2	81	1	US-08-443-407-88
26	15.8	75.2	81	5	PCT-US95-03600-232
27	15.8	75.2	1433	1	US-07-968-971A-11

c 28	15.8	75.2	1433	1	US-08-383-756-5
c 29	15.8	75.2	1433	1	US-08-424-406-2
c 30	15.8	75.2	1433	1	US-08-464-523B-8
c 31	15.8	75.2	1433	2	US-08-460-898-5
c 32	15.2	72.4	262	4	US-09-437-457-13
c 33	15.2	72.4	1618	1	US-08-410-540-1
c 34	15.2	72.4	1641	4	US-08-659-254-19
c 35	15.2	72.4	2811	4	US-08-482-073-3
c 36	15.2	72.4	2813	2	US-08-344-155C-99
c 37	15.2	72.4	2813	4	US-09-009-490A-90
c 38	15.2	72.4	3080	4	US-08-482-073-4
c 39	15.2	72.4	4016	1	US-08-410-540-3
c 40	14.8	70.5	1925	4	US-08-894-324A-1
c 41	14.8	70.5	2778	1	US-08-202-054-1
c 42	14.8	70.5	2778	1	US-08-446-923-1
c 43	14.6	69.5	3390	4	US-09-550-338-1
c 44	14.4	68.6	1498	1	US-07-965-668A-1
c 45	14.4	68.6	1498	2	US-08-950-433-1

ALIGNMENTS

RESULT 1  
US-08-064-121-2  
; Sequence 2, Application US/08064121  
; Patent No. 5641664  
; GENERAL INFORMATION:  
; APPLICANT: D'HALLUIN, Kathleen  
; APPLICANT: GOBEL, Elke  
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/064,121  
; FILING DATE: 24-MAY-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 90403332.1  
; FILING DATE: 23-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91401888.2  
; FILING DATE: 08-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crane-Feury, Sharon E  
; REGISTRATION NUMBER: 36,113  
; REFERENCE/DOCKET NUMBER: 010830-043  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1186 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORGANISM: probe

```

;
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycinase
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-08-064-121-2

```

```

Query Match 92.4%; Score 19.4; DB 1; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 gcttgactataactctgac 21
|||||
Db 1113 GCTTGGACTATAATACCTGAC 1133

```

```

RESULT 2
US-08-478-015-2
; Sequence 2, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:

```

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycinase phosphotransferase gc
; US-08-478-015-2

```

```

Query Match 92.4%; Score 19.4; DB 1; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 gcttgactataactctgac 21
|||||
Db 1113 GCTTGGACTATAATACCTGAC 1133

```

```

RESULT 3
US-08-475-975-2
; Sequence 2, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:

```

```
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-08-475-975-2

Query Match 92.4%; Score 19.4; DB 3; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataatacttgac 21
|||||
Db 1113 GCTTGGACTATAAATACCTGAC 1133

RESULT 4
US-09-084-889-2
; Sequence 2, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: Gobel, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/064,121
; FILING DATE:
; PRIOR APPLICATION DATA: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-09-084-889-2

Query Match 92.4%; Score 19.4; DB 3; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataatacttgac 21
|||||
Db 1113 GCTTGGACTATAAATACCTGAC 1133

RESULT 5
US-08-894-440-2
; Sequence 2, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
; OTHER INFORMATION: fragment of pTS88
```

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(35)
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36)..(694)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain
; OTHER INFORMATION: CM1841 (P35S)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (695)..(967)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (968)..(1287)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1288)..(1303)
; OTHER INFORMATION: polylinker of pGEM2
;
; US-08-894-440-2
```

```
Query Match 92.4%; Score 19.4; DB 3; Length 1303;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataactctgac 21
|||||
Db 1178 gcttgactataactctgac 1198
```

```
RESULT 6
US-09-080-625-3
; Sequence 3, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; APPLICANT: Spencer, T. Michael
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
;
; TOPOLOGY: linear
; US-09-080-625-3

Query Match 92.4%; Score 19.4; DB 4; Length 3153;
Best Local Similarity 95.2%; Pred. No. 0.32;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataactctgac 21
|||||
Db 2759 GCTTGGACTATAATACCTGAC 2779

RESULT 7
US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,104
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
;
; US-08-453-104-23

Query Match 92.4%; Score 19.4; DB 1; Length 3200;
Best Local Similarity 95.2%; Pred. No. 0.32;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 gcttgactataacttgac 21  
|||||  
Db 3032 GCTGGACTATAACTGAC 3052

## RESULT 8

US-08-694-824-23  
; Sequence 23, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 92.4%; Score 19.4; DB 2; Length 3200;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21  
|||||  
Db 3032 GCTGGACTATAACTGAC 3052

## RESULT 9

US-08-453-104-22  
; Sequence 22, Application US/08453104

; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 92.4%; Score 19.4; DB 1; Length 3201;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21  
|||||  
Db 3105 GCTGGACTATAACTGAC 3125

## RESULT 10

US-08-694-824-22  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

;; TITLE OF INVENTION: IN PLANT CELLS  
;; NUMBER OF SEQUENCES: 23  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: George Mason Bldg., Washington & Prince Sts.  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22133-1404

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.125  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/694,824

;; FILING DATE: 09-AUG-1996  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/937,869  
;; FILING DATE: 16-DEC-1992  
;; APPLICATION NUMBER: GB 90401055.0  
;; FILING DATE: 18-APR-1990

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Rea, Teresa S  
;; REGISTRATION NUMBER: 30,427  
;; REFERENCE/DOCKET NUMBER: 010830-032  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620

;; INFORMATION FOR SEQ ID NO: 22:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3201 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)

;; FEATURE:

;; NAME/KEY: misc.feature

;; LOCATION: 2151..2155

;; OTHER INFORMATION: /note= "Nucleotides 2151-2155

;; US-08-694-824-22

Query Match 92.4%; Score 19.4; DB 2; Length 3201;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21  
|||||  
Db 3105 GCTTGACTATACTGAC 3125

RESULT 11  
US-09-080-625-2  
;; Sequence 2, Application US/09080625  
;; Patent No. 6307123  
;; GENERAL INFORMATION:  
;; APPLICANT: Kriz, Alan L.  
;; APPLICANT: Spencer, T. Michael  
;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Arnold White & Durkee  
;; STREET: P.O. Box 4433  
;; CITY: Houston  
;; STATE: TX  
;; COUNTRY: USA  
;; ZIP: 77210-4433

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/080,625  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hanson, Robert E.  
;; REGISTRATION NUMBER: P-42,628  
;; REFERENCE/DOCKET NUMBER: DEKM:161  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (512) 418-3000  
;; TELEFAX: (512) 474-7577

;; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3336 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear

;; US-09-080-625-2

Query Match 92.4%; Score 19.4; DB 4; Length 3336;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21  
|||||

Db 2942 GCTTGACTATACTGAC 2962

RESULT 12  
US-09-080-625-5

;; Sequence 5, Application US/09080625

;; Patent No. 6307123

;; GENERAL INFORMATION:  
;; APPLICANT: Kriz, Alan L.  
;; APPLICANT: Spencer, T. Michael  
;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Arnold White & Durkee  
;; STREET: P.O. Box 4433  
;; CITY: Houston  
;; STATE: TX  
;; COUNTRY: USA  
;; ZIP: 77210-4433

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/080,625  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hanson, Robert E.  
;; REGISTRATION NUMBER: P-42,628  
;; REFERENCE/DOCKET NUMBER: DEKM:161  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (512) 418-3000  
;; TELEFAX: (512) 474-7577

;; INFORMATION FOR SEQ ID NO: 5:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3694 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear

;; US-09-080-625-5



Query Match 92.4%; Score 19.4; DB 4; Length 3694;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttggaactataacttgac 21  
|||||  
DB 3300 GCTTGGACTATAATACCTGAC 3320

RESULT 13  
US-09-080-625-4  
; Sequence 4, Application US/09080625  
; Patent No. 6307123  
; GENERAL INFORMATION:  
; APPLICANT: Kriz, Alan L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/09/080, 625  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, Robert E.  
; REGISTRATION NUMBER: P-42, 628  
; REFERENCE/DOCKET NUMBER: DEKM:161  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3877 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-080-625-4

Query Match 92.4%; Score 19.4; DB 4; Length 3877;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttggaactataacttgac 21  
|||||  
DB 3483 GCTTGGACTATAATACCTGAC 3503

RESULT 14  
US-08-817-188-1/c  
; Sequence 1, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; CURRENT FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
; EARLIER FILING DATE: 1996-07-31

; EARLIER APPLICATION NUMBER: EP 95401844.6  
; EARLIER FILING DATE: 1995-08-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 1  
; LENGTH: 4946  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; OTHER INFORMATION: plasmid pTHW107  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((1)..(25))  
; OTHER INFORMATION: T-DNA right border (RB)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((97)..(330))  
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
; OTHER INFORMATION: T-DNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((331)..(882))  
; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
; OTHER INFORMATION: transferase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((883)..(2608))  
; OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2658)..(3031))  
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
; OTHER INFORMATION: gene of Agrobacterium T-DNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((3032)..(3367))  
; OTHER INFORMATION: barnase: region coding for barnase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((3368)..(4876))  
; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana  
; OTHER INFORMATION: tabacum  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((4922)..(4946))  
; OTHER INFORMATION: LB: T-DNA left border  
US-08-817-188-1

Query Match 92.4%; Score 19.4; DB 3; Length 4946;  
Best Local Similarity 95.2%; Pred. No. 0.34;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttggaactataacttgac 21  
|||||  
DB 163 GCTTGGACTATAATACCTGAC 143

RESULT 15  
US-08-817-188-5/c  
; Sequence 5, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; CURRENT FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366

```

; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-LSI gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5067)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
;
; US-08-817-188-5
```

```

Query Match          92.4%; Score 19.4; DB 3; Length 5560;
Best Local Similarity 95.2%; Pred. No. 0.34;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttggactataatacttgac 21
   |||||
Db 150 GCTTGGACTATAATACCTGAC 130
```

Search completed: February 15, 2002, 19:09:18  
Job time: 14732 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:04:17 ; Search time 9904.61 Seconds  
(without alignments)  
22.783 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21  
Sequence: 1 gcttgactataactgac 21

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST.\*
- 1: em\_estfun:\*\*
  - 2: em\_esthum:\*\*
  - 3: em\_estin:\*\*
  - 4: em\_estom:\*\*
  - 5: em\_estpl:\*\*
  - 6: em\_estba:\*\*
  - 7: em\_estro:\*\*
  - 8: em\_estov:\*\*
  - 9: em\_htc:\*\*
  - 10: gb\_est1:\*\*
  - 11: gb\_est2:\*\*
  - 12: gb\_htc:\*\*
  - 13: gb\_gss:\*\*
  - 14: em\_gss\_fun:\*\*
  - 15: em\_gss\_hum:\*\*
  - 16: em\_gss\_inv:\*\*
  - 17: em\_gss\_pln:\*\*
  - 18: em\_gss\_pro:\*\*
  - 19: em\_gss\_rod:\*\*
  - 20: em\_gss\_vrt:\*\*
  - 21: em\_gss\_other:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.4	82.9	344	AQ117151	AQ117151 HS_2188_A
2	17.4	82.9	633	13	AZ464390 1M0273D15
3	16.8	80.0	291	10	AV035892 AV035892
4	16.8	80.0	415	13	AQ172470 HS_3190_B
C 5	16.8	80.0	449	10	AI675738 wc40h04.x
6	16.8	80.0	449	11	BE802728 sr44g04.y
C 7	16.8	80.0	470	13	AZ133172 OSJNB010
8	16.8	80.0	483	10	AA310103 EST180941
9	16.8	80.0	487	13	B74437 CIT-HSP-202
C 10	16.8	80.0	505	13	AQ666336 HS_5374_A
11	16.8	80.0	537	13	AQ836967 HS_5353_A
12	16.8	80.0	600	11	BE846195 232148 BA

C 13	16.8	80.0	854	13	BH132866
C 14	16.8	80.0	968	13	AZ670679 ENTHD34TF
15	16.4	78.1	301	10	BB103479
16	16.4	78.1	302	10	AA407887
C 17	16.4	78.1	305	10	BB117588
18	16.4	78.1	309	10	BB235252
19	16.4	78.1	310	10	AU021901
C 20	16.4	78.1	353	10	AU021997
C 21	16.4	78.1	376	10	AA823300
C 22	16.4	78.1	405	10	AA606880
C 23	16.4	78.1	408	10	AU021898
24	16.4	78.1	432	10	AA763279
25	16.4	78.1	453	13	AQ236085
26	16.4	78.1	470	10	AU017121
C 27	16.4	78.1	493	10	AA690887
28	16.4	78.1	503	10	AU042554
29	16.4	78.1	514	10	AI315372
C 30	16.4	78.1	576	10	BE325283
31	16.4	78.1	605	10	AW109828
C 32	16.4	78.1	613	13	AZ411070
C 33	16.4	78.1	667	13	AQ198178
C 34	16.4	78.1	678	13	AQ201499
35	16.4	78.1	863	11	BF978014
36	16.2	77.1	280	10	BB337839
C 37	16.2	77.1	305	10	AA013302
C 38	16.2	77.1	411	10	AA554565
C 39	16.2	77.1	412	11	BF841371
C 40	16.2	77.1	423	13	AZ273622
41	16.2	77.1	441	11	H04932
42	16.2	77.1	445	13	AQ710257
43	16.2	77.1	458	13	AQ493489
44	16.2	77.1	468	10	BE246449
45	16.2	77.1	485	13	AG023449

ALIGNMENTS

RESULT 1  
AQ117151/c

LOCUS HS\_2188\_A1\_D04\_MF CIR Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2188 Col=7 Row=G, DNA sequence.

DEFINITION HS\_2188\_A1\_D04\_MF CIR Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2188 Col=7 Row=G, DNA sequence.

ACCESSION AQ117151

VERSION AQ117151.1 GI:3494942

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 344)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 2188 row: G column: 7  
Class: BAC ends  
High quality sequence stop: 344.  
location/Qualifiers  
1. 344  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="Plate=2188 Col=7 Row=G"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT 105 a 53 c 85 g 101 t  
ORIGIN

Query Match 82.9%; Score 17.4; DB 13; Length 344;  
Best Local Similarity 94.7%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ctgggactataacttga 20  
|||||||

Db 167 CATGGACTATAAATCTTGA 149

## RESULT 2

AZ464390 633 bp DNA GSS 04-OCT-2000  
LOCUS  
DEFINITION  
1M0273D15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0273D15 R, DNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AZ464390.1 GI:106222515

GSS.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 633)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

Plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0273 row: D column: 15

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 633.

Location/Qualifiers

1. .633

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0273D15"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid p1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 183 a 133 c 136 g 181 t  
ORIGIN

Query Match 82.9%; Score 17.4; DB 13; Length 633;  
Best Local Similarity 94.7%; Pred. No. 2.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ctgggactataacttga 20  
|||||||

Db 473 CTTGGACTAAATCTTGA 491

## RESULT 3

AV035892 291 bp mRNA EST 22-NOV-1999  
LOCUS  
DEFINITION  
AV035892 Mus musculus adult C57BL/6J placenta Mus musculus cDNA  
clone 1600016H12, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AV035892.1 GI:4855557

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 291)

Carinci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,

Akaira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara

,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,

Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuna,H., Oda,H., Owa,C.,

Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara

,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomioka,N.,

Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,

Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

Location/Qualifiers

1. .291

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="1600016H12"

/clone\_lib="Mus musculus adult C57BL/6J placenta"

/sex="female"

/tissue\_type="placenta"

/dev\_stage="adult"

68 a 56 c 66 g 101 t

BASE COUNT

ORIGIN

Query Match 80.0%; Score 16.8; DB 10; Length 291;  
Best Local Similarity 90.0%; Pred. No. 4.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	2	cttggactataacttgac 21 	415 bp	DNA	17-OCT-1998
Db	13	CTTGGACTATATACCTGGAC 32			
RESULT	4				
QY	172470	AQ172470	415 bp	DNA	GSS
LOCUS		HS_3190_B2_H07_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3190 Col=14 Row=P, DNA sequence.			
DEFINITION		AQ172470			
ACCESSION		AQ172470			
VERSION		AQ172470.1	GI:3569837		
KEYWORDS		GSS.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS		1 (bases 1 to 415)			
TITLE		Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.			
JOURNAL		Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome			
MEDLINE		Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)			
COMMENT		99380589 Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3190 row: P column: 14 Class: BAC ends High quality sequence stop: 415.			
FEATURES		source			
		1. .415			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone="Plate=3190 Col=14 Row=P"			
		/clone_lib="CIT Approved Human Genomic Sperm Library D"			
		/sex="male"			
		/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"			
BASE COUNT		77 a 117 c 69 g 151 t			
ORIGIN					
		Query Match 80.0%; Score 16.8; DB 13; Length 415;			
		Best Local Similarity 90.0%; Pred. No. 4.8e+02;			
		Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	2	cttggactataacttgac 21 	449 bp	mRNA	EST
Db	27	CTTGGATATATAGCTTGAC 46			
RESULT	5				
QY	5738/c	AI675738/c	449 bp	mRNA	EST
LOCUS		WC40h04. x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321143 3' similar to gb:M73255_rnal VASCULAR CELL ADHESION PROTEIN 1			
DEFINITION		AI675738			
ACCESSION		AI675738			
VERSION		AI675738.1	GI:4876218		
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
		1 (bases 1 to 449)			
AUTHORS		1 (bases 1 to 449)			
TITLE		Shoemaker, R., Kelm, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.			
JOURNAL		Public Soybean EST Project			
COMMENT		Unpublished (1999) Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project			
AUTHORS		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert length: 567 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 411.			
JOURNAL		Location/Qualifiers			
COMMENT		1. .449			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone="IMAGE:2321143"			
		/clone_lib="NCI_CGAP_Pr28"			
		/sex="male"			
		/dev_stage="adult"			
		/lab_host="DH10B"			
		/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaudo."			
BASE COUNT		178 a 77 c 63 g 131 t			
ORIGIN					
		Query Match 80.0%; Score 16.8; DB 10; Length 449;			
		Best Local Similarity 90.0%; Pred. No. 4.8e+02;			
		Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1	gcttgactataacttgac 20 	449 bp	mRNA	EST
Db	419	GCTTGACTATATATTTA 400			
RESULT	6				
QY	2728/c	BE802728/c	449 bp	mRNA	EST
LOCUS		Sr44g04.y1 Gm-cl051 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl051-943 5', mRNA sequence.			
DEFINITION		BE802728			
ACCESSION		BE802728			
VERSION		BE802728.1	GI:10233840		
KEYWORDS		EST.			
SOURCE		soybean.			
ORGANISM		Glycine max			
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.			
AUTHORS		1 (bases 1 to 449)			
TITLE		Shoemaker, R., Kelm, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.			
JOURNAL		Public Soybean EST Project			
COMMENT		Unpublished (1999) Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project			

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 High quality sequence stop: 402.

## FEATURES

source

1..449

/organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1051-943"  
 /clone\_lib="Gm-c1051"  
 /tissue\_type="floral meristematic mRNA"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:  
 XhoI; The cDNA library was constructed from floral  
 meristematic mRNA provided by Dr. Halina Knap of Clemson  
 University. Complementary DNA was synthesized from mRNA  
 using a primer consisting of a poly(dT) sequence with a  
 XhoI restriction site. EcoRI adapters were ligated to the  
 blunt-ended cDNA fragments followed by XhoI digestion. The  
 cDNA fragments were directionally cloned into the  
 EcoRI-XhoI restriction site of the pBluescript vector. The  
 ligated cDNA fragments were transformed into DH10B host  
 cells (GibcoBRL). This library was constructed in the  
 laboratory of Dr. Randy Shoemaker."

BASE COUNT

ORIGIN

149 a 73 c 80 g 147 t

Query Match

80.0%; Score 16.8; DB 11; Length 449;

Best Local Similarity 90.0%; Pred. No. 4.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ctgggactataacttgac 21

|||||

Db 358 CTGGACTATAAAGTTGAC 339

RESULT 7

AZ133172

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzoaceae; Oryza.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

A BAC End Sequencing Framework to Sequence the Rice Genome  
 Unpublished (1998)  
 Contact: Wing RA  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: GGAACAGCATGACCATG  
 Class: BAC ends  
 High quality sequence start: 101  
 High quality sequence stop: 368.  
 Location/Qualifiers

FEATURES

source

1..470

/organism="Oryza sativa"

/strain="Japonica"

/cultivar="Nipponbare"

/db\_xref="taxon:4330"

/clone="OSUNBB0108P05r"

/clone\_lib="CUGI Rice BAC Library (ECORI)"

/tissue\_type="Leaf"

/lab\_host="E. coli DH10B"

/note="Vector: pBACIndigo; Site\_1: EcoRI; Site\_2: EcoRI;  
 Rice is the most important food crop in the world. Half of  
 the world population, especially those inhabiting highly  
 populated areas of the humid tropics and subtropics, rely  
 on rice as their primary source of carbohydrate.  
 Monocotyledonous rice is a diploid plant (2n=24) with a  
 haploid genome equivalent of 431 Mbp (Arumuganathan and  
 Earle, 1991). The relatively small genome of rice, three  
 times larger than that of Arabidopsis, makes it suitable  
 for genomic studies. In order to facilitate positional  
 cloning, physical mapping and genome sequencing of rice,  
 we have constructed a BAC library from Oryza sativa.  
 Nipponbare variety using EcoRI as the cloning enzyme. The  
 library contains 55,296 clones with an average insert size  
 of 121 kb providing approximately 15 haploid genome  
 equivalents. The deep coverage allows the isolation a  
 particular sequence with a probability of 99.9 %. Three  
 high density filters, each containing 18,432 clones  
 (doubly spotted), represent the whole library for colony  
 screening and can be requested from the Clemson University  
 BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT

ORIGIN

163 a 75 c 94 g 131 t

Query Match

80.0%; Score 16.8; DB 13; Length 470;

Best Local Similarity 90.0%; Pred. No. 4.9e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 20

|||||

Db 277 GCTGGGCTCAATACTTGA 296

RESULT 8

AA310103

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Human.  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 483)  
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult  
 C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White  
 O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,  
 Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald  
 L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,  
 Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,  
 Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
 Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,  
 Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,  
 Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,  
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
 Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,  
 Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,  
 Kunsch, C., Hungjun, J., Li, H., Meissner, P., Olsen, H., Raymond, L.,  
 Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon  
 M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and  
 Venter, J.C.

**TITLE** Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
**JOURNAL** Nature 377 (6547 Suppl), 3-174 (1995)  
**MEDLINE** 96026280  
**COMMENT** Other ESTs: THC122434  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org

For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

#### FEATURES

Location/Qualifiers  
 1. .483

/organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):156165"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Jurkat T-cells v"  
 /cell\_type="T-lymphocyte"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

**BASE COUNT** 113 a 142 c 136 g 85 t 7 others

#### ORIGIN

**Query Match** 80.0%; Score 16.8; DB 10; Length 483;  
**Best Local Similarity** 85.7%; Pred. No. 4.9e+02;  
**Matches** 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY** 1 gcttgactataactctgac 21  
 ||||| ||| |||||

**Db** 123 GCTTGGACGAAATCTTGAC 143

#### RESULT

9

#### LOCUS

**B74437** 487 bp DNA GSS 24-OCT-1998  
**DEFINITION** CIT-HSP-2028F18.TF CIT-HSP Homo sapiens genomic clone 2028F18, DNA  
 sequence.

#### ACCESSION

**B74437**

#### VERSION

**B74437.1** GI:2770124

#### KEYWORDS

**GSS.**

#### SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 487)

Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden

,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.

and Venter,J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map

Building

Unpublished (1997)

Other GSSs: CIT-HSP-2028F18.TR

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)

Seq primer: M13-21

Class: BAC ends.

Location/Qualifiers

1. .487

/organism="Homo sapiens"

/db\_xref="GDB:7048076"  
 /db\_xref="taxon:9606"  
 /clone="2028F18"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /note="vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
 HindIII"

**BASE COUNT** 160 a 101 c 100 g 126 t

#### ORIGIN

**Query Match** 80.0%; Score 16.8; DB 13; Length 487;  
**Best Local Similarity** 90.0%; Pred. No. 4.9e+02;  
**Matches** 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 2 cttggactataactctgac 21  
 ||||| ||||| |||||

**Db** 57 CTTGGACTGTAATACTTGTC 76

#### RESULT

10

#### LOCUS

**AQ666336/c**

#### DEFINITION

HS\_5374\_AL\_H01.T7A RPCI-11 Human Male BAC Library Homo sapiens

genomic clone Plate=950 Col=1 Row=O, DNA sequence.

#### ACCESSION

**AQ666336**

#### VERSION

**AQ666336.1** GI:5174104

#### KEYWORDS

**GSS.**

#### SOURCE

human.

#### ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 505)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))

or from Resear h Genetics (info@resgen.com). BAC end Web Server:

<http://www.htsc.washington.edu>

Plate: 950 row: 0 column: 1

Seq primer: T7

Class: BAC ends

High quality sequence stop: 505.

Location/Qualifiers

1. .505

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=950 Col=1 Row=O"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the

pBACe3 6 vector at EcoRI sites"

141 a 121 c 85 g 154 t 4 others

**BASE COUNT**

**ORIGIN**

```

Query Match      80.0%; Score 16.8; DB 13; Length 505;
Best Local Similarity 90.0%; Pred. No. 4.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactttgac 21
||||| ||| ||||| |||
Db 500 CTGGAGTATATATCTTGC 481

RESULT 11
A0836967
LOCUS      A0836967      537 bp      DNA      GSS      30-AUG-1999
DEFINITION HS_4553_Al_D09_T7A CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=4553 Col=17 Row=G, DNA sequence.
ACCESSION  A0836967
VERSION     A0836967.1 GI:5806929
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 537)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380599
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 4553 row: G column: 17
Seq primer: T7
Class: BAC ends
High quality sequence stop: 537.
Location/Qualifiers
1..537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=4553 Col=17 Row=G"
/sex="male"
/note="Organ: sperm; Vector: pBelorAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT  154 a 128 c 115 g 137 t 3 others
ORIGIN
cttggactataactttgac 21
||||| ||| ||||| |||

Query Match      80.0%; Score 16.8; DB 13; Length 537;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactttgac 21
||||| ||| ||||| |||
Db 438 CTGGACTAGTACTTTAC 457

RESULT 12
B846195
LOCUS      B846195      600 bp      mRNA      EST      25-SEP-2000
DEFINITION B23148 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  B846195
VERSION     B846195.1 GI:10283019
KEYWORDS    EST.
SOURCE      cow.

```

```

ORGANISM    Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 600)
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
Wells,K.D.
Mapping of Expressed Sequence Tags from a normalized bovine mammary
gland cDNA library
Unpublished (2000)
Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 112 row: L column: 9
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..600
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

BASE COUNT  150 a 140 c 112 g 198 t
ORIGIN
cttggactataactttgac 21
||||| ||| ||||| |||
Db 532 CTGGACTCTAATCCTTGAC 551

Query Match      80.0%; Score 16.8; DB 11; Length 600;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactttgac 21
||||| ||| ||||| |||
Db 532 CTGGACTCTAATCCTTGAC 551

RESULT 13
BH132866/c
LOCUS      BH132866      854 bp      DNA      GSS      07-AUG-2001
DEFINITION ENTNI55TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic. DNA sequence.
ACCESSION  BH132866
VERSION     BH132866.1 GI:15091927
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
ORGANISM    Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 854)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: entae@igr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library

```



Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 18  
 High quality sequence stop: 742.

## FEATURES

source

Location/Qualifiers  
 1..854  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /note="Vector: PHOS1; Site\_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT

251 a 100 c 95 g 408 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 854;  
 Best Local Similarity 90.0%; Pred. No. 5.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcttgactataacttga 20

|||||

Db 143 GCTTGGACTAAATATTGA 124

RESULT 14

AZ670679/c

LOCUS

DEFINITION ENT670679 Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION AZ670679

VERSION 1

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 968)

AUTHORS Loftus, B., Van Aken, S. and Fraser, C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library

COMMENT Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: enta@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 94

High quality sequence stop: 406.

Location/Qualifiers

1..968

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db\_xref="taxon:5759"

/clone\_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOS1; Site\_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

FEATURES

source

Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT

254 a 127 c 210 g 377 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 968;  
 Best Local Similarity 90.0%; Pred. No. 5.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcttgactataacttga 20

|||||

Db 363 GCTTGGACTAAATATTGA 344

RESULT 15

BB103479

LOCUS

DEFINITION BB103479 RIKEN full-length enriched, 12 days embryo, embryonic body between diaphragm region and neck Mus musculus cDNA clone 9430087K10 3', mRNA sequence.

ACCESSION BB103479

VERSION BB103479.1

KEYWORDS GI:8756047

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 301)

AUTHORS Konno, H., Alzawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci

P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, K., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Toninaga, N., Toya, T., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yamakawa, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp

URL: http://genome.gsc.riken.go.jp/

N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
 source  
 Location/Qualifiers  
 1. .301  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="943087K10"  
 /clone\_lib="RIKEN full-length enriched, 12 days embryo,  
 embryonic body between diaphragm region and neck"  
 /tissue\_type="embryonic body between diaphragm region and  
 neck"  
 /dev\_stage="12 days embryo"  
 /lab\_host="DH10B"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGAGATCCCAAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. cDNA went through one round of normalization  
 to Rot = 20.0 and subtraction to Rot = 370.0. Second  
 strand cDNA was prepared with the primer adapter of  
 sequence [5' GAGAGAGAGATTCGAGTTAATAATTAATCCCCCCCCC  
 3']. cDNA was cleaved with XhoI and BamHI."

BASE COUNT 75 a 54 c 48 g 124 t  
 ORIGIN

Query Match 78.1%; Score 16.4; DB 10; Length 301;  
 Best Local Similarity 94.4%; Pred. No. 7.1e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ttggactataacttga 20  
 ||||| |||||  
 Db 96 TTGGACTTTAATACTTGA 113

Search completed: February 15, 2002, 18:04:21  
 Job time: 20970 sec



```
sig_peptide
17. .85
/note="cruciferin signal peptide"
```

Unclassified.  
REFERENCE 1 (bases 1 to 3113)

sig\_pe

```
misc_feature      2015. .2920  
                  /note="polyA signal"  
misc_feature      2989. .2994  
                  /note="polyA signal"
```

BASE COUNT	50.1	a	000	C	300	g	001	C
ORIGIN								

	22,	conservative	0,	mismatches	0,	indels	0,	gaps	0,
Qy	1	aacgagtgtagcagaccgc	22						

```

DO      750  AACGAGGAGCAGCAGAACCCACAC  811
      RESULT      7

```

LOCUS	DEFINITION	NCBI	3198 bp	DNA	PLN
	B.napus BnCl gene for cruciferin storage protein.				04-APR-1995
ACCESSION		X59294			
VERSION		X59294.1	GI:17790		
KEYWORDS		cruciferin; cruciferin storage protein.			
SOURCE		rape.			

**ORGANISM**  
*Brassica napus*  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE  
Rosidae: eurosoids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 3198)  
AUTHORS  
Breen, J. P. and Crouch, M. L.  
TITLE  
Molecular analysis of a cruciferin storage protein gene family of  
Brassica napus

JOURNAL Plant Mol. Biol. 19 (6), 1049-1055 (1992)  
MEDLINE 92379259  
REFERENCE 2 (bases 1 to 3198)  
AUTHORS Breen, J.P.  
TITLE Direct Submission  
JOURNAL Submitted (24-APR-1991) Green, J.P., Jordan Hall, Indiana

COMMENT	Submitted: 2007-07-17; Organism: <i>Salmonella enterica</i>
FEATURES	See also M16860 & X59295 (for Bnc2 gene).
SOURCE	Location/Qualifiers 1..3198 /organism="Salmonella enterica"

```

/organism="Brassicaceae"
/strain="C.v. toweri"
/db_xref="taxon:3708"
/cell_line="ED8767"
/clone_lib="Ch4a"
/clone="lambda BnCl1"
join(<709..991,1220..1584,2069..2488,2653..>3057)
/gene="BnCl1"
mRNA

```

```

gene
709..3057
/genes="BnCl1"
join(709..991,1220..1584,2069..2488,2653..3057)
/genes="BnCl1"
/codon_start=1
/product="cruciferin storage protein"
/protein_id="CAA41984.1"
/db_xref="GI:762919"

```

```

/db_xref="SWISS-PROT:P33523"
/translation="MARLSGLSFLALLIFLHGSTAQQPPNECQLDQLNALPESHVLV
KAEAGRIEYMDHAPQLSGSGSFVRYIIESKGLYLPSEFTAKLFGVAKEGELMGVRV
VPCDTESSVFPQSGSGSQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
GFGMHQKVEHIFTGDTIATHPCVAQWYNDGNQPLIVSVLDSHGONOLDLRNPRF
YLAGNPPQGVWTEGREQPQKNLNGFTPEVLAKAFKIDVRTAQQLOONQDNRGNII
RVGPESVIRPLPQRSQPQETVNWGLEETICSARCDNLDDPDSADVYKFPQLGISTL
NSYDPLIRFLRLSALSGIRQNALNANAVLYVTDGEAHVYVQVNDNRDVRV
DGVSGOGLLISIQFGFSVVKRATSEQRTFEFTNANAQNTLAGRTSVLRGLPLFVIV
SNGCYOILFEARVRVKNTFTFTTLTHSSGSPASYGGRPKADA"

```

exon	<709..991
	/gene="BnCl"
	/number=1
intron	992..1219

```

/feature="BnCl"
/number=1
1220..1584
/feature="BnCl"
/number=2
1585..2068
/feature="BnCl"
/number=2
2069..2488
/feature="BnCl"
/number=3
2489..2652
/feature="BnCl"
/number=3
2654..>3057
/feature="BnCl"
/number=4
3102..3107
polyA-signal
polyA-signal
3176..3181
BASE COUNT 992 a 690 c 600 g 916 t
ORIGIN

```

Query Match 100.0%; Score 22; DB 8; Length 3198;  
 Best Local Similarity 100.0%; Pred. No. 0.1;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 aacgagtgtcagctagaccagc 22
|||||
Db 790 AACGAGTGTCTAGTACCAGC 811

```

## RESULT 8

AL139402 Human DNA sequence from clone RP4-808F24 on chromosome Xp11.23-11.4. Contains STSs and GSSs, complete sequence.  
 AL139402 AL139402.11 GI:9944136 human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 I (bases 1 to 10959)  
 Direct Submission  
 Submitted (24-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk

## COMMENT

On Aug 29, 2000 this sequence version replaced gi:9926454.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>  
 RP4-808F24 is from the library RPCI-4 constructed at the Roswell

Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>  
 VECTOR: pCYPAC2  
 IMPORTANT: This sequence is not the entire insert of clone RP4-808F24. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP4-774c10 is at 10860 in this sequence.  
 The true right end of clone RP6-227L5 is at 100 in this sequence.

## FEATURES

## source

```

1..10959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11.23-11.4"
/clone="RP4-808F24"
/clone_lib="RPCI-4"
/complement(1..87)
/note="match: STS: Em:HS227L5T"
2314..2355
/note="21 copies 2 mer gg 76% conserved"
2339..2342
/note="unidirectional dGTP"
/complement(2797..3221)
/note="match: GSS: Em:B83505"
5384..5631
/note="L1MB8 repeat: matches 5924..6175 of consensus"
5682..5791
/note="L1MB8 repeat: matches 5654..5773 of consensus"
6792..6940
/note="LTR37A repeat: matches 268..420 of consensus"
6941..6978
/note="19 copies 2 mer ta 92% conserved"
6980..7218
/note="LTR37A repeat: matches 47..282 of consensus"
7440..7486
/note="LTR37A repeat: matches 1..47 of consensus"
8540..8922
/note="MSTA repeat: matches 1..426 of consensus"
/complement(8828..9278)
/note="match: GSS: Em:AQ23257"
/complement(8948..9266)
/note="match: GSS: Em:B53441"
9300..9420
/note="MER5B repeat: matches 1..128 of consensus"
/complement(9616..10040)
/note="match: GSS: Em:AQ369714"
10048..10487
/note="L1MB3 repeat: matches 5455..5901 of consensus"
10509..10820
/note="LTR26 repeat: matches 291..600 of consensus"
10847..10958
/note="LTR26 repeat: matches 201..320 of consensus"
BASE COUNT 3380 a 2058 c 2220 g 3301 t
ORIGIN

```

Query Match 88.2%; Score 19.4; DB 9; Length 10959;  
 Best Local Similarity 95.2%; Pred. No. 3.7;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 aacgagtgtcagctagaccag 21
|||||
Db 7197 AAGGAGTGTCTAGTACCAG 7217

```

## RESULT 9

ATCRAL LOCUS 2346 bp DNA 18-MAY-1995  
 DEFINITION Arabidopsis CRL gene for 12S seed storage protein.  
 ACCESSION X14312  
 VERSION X14312.1 GI:16231  
 KEYWORDS CRL gene; seed storage protein; storage protein.  
 SOURCE thale cress.

ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 2346)  
AUTHORS Pang,P.P., Pruitt,R.E. and Meyerowitz,E.M.  
TITLE Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana  
JOURNAL Plant Mol. Biol. 11, 805-820 (1988)  
FEATURES  
source 1. .2346  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/cultivar="Landsberg erecta"  
/db\_xref="taxon:3702"  
/clone\_lib="lambda EMBL4"  
promoter 138. .147  
/note="TATA-box"  
exon 196. .496  
/number=1  
CDS join(196. .496,623. .888,985. .1431,1547. .1951)  
/product="12S seed storage protein"  
/protein\_id="CAA32493.1"  
/db\_xref="GI:808936"  
/db\_xref="SWISS-PROT:P15455"  
/translation="MARVSLLSFCLTLFLFHGYAAQQGQGGQFFNECQLDQLNAL  
EPSHVLKSEAGRIEVDHHPALQRCGVSFARYIIIESKGLYLPSFNTAKLSFVAKGR  
GLMGKIPGCAETFDQSFQRFEGQSGQSRFDMHVKVEHIRSGDTIATTPGVAQM  
FYNDGQGPLVIVSFDLASHQNDLRNPRFYLAGNNPQGVWLGREQPQKNIFNG  
FGPEVTAQALKIDLTQAQQLQNDNRNIVRVGGPFGVIRPRLRGQPEEEEEER  
HGRNGLEETICSARCTNDLDDPSRADVYKPOLGYISTLSNYDLPLIRLIRLSALRG  
SIRQNAVLPMWNAANAILIETDGEAQIOIVNDNGNRVFDGVSGQLIAPQGFVS  
VKRATSNRFQWVEFTNNAQAINTLAGRTSVLRGLPLEVITNGFQISPEEARVKNT  
LETLTHSSGPASVGRPRVAA"  
intron 497. .622  
/number=1  
exon 623. .888  
/number=2  
intron 889. .984  
/number=2  
exon 985. .1431  
/number=3  
intron 1432. .1546  
/number=3  
exon 1547. .1951  
/number=4  
polyA\_signal 2015. .2020  
polyA\_signal 2085. .2090  
/notes="alternative"  
BASE COUNT 714 a 528 c 472 g 632 t  
ORIGIN  
Query Match 85.5%; Score 18.8; DB 8; Length 2346;  
Best Local Similarity 90.9%; Pred. No. 7.8;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 aacgagtgcagctagaccagc 22  
||||||| ||||| |||||  
Db 295 AACGAGTGCACCTGACCAGC 316  
RESULT 10  
LOCUS ATHCRA1AA 2346 bp DNA PLN 27-APR-1993  
DEFINITION A.thaliana 12S storage protein CRA1 gene, exons 1-4.  
ACCESSION M37247  
VERSION M37247.1 GI:166675  
KEYWORDS 12S storage protein.  
SOURCE A.thaliana, cDNA to mRNA.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 2346)  
AUTHORS Pang,P.P., Pruitt,R.E. and Meyerowitz,E.M.  
TITLE Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana  
JOURNAL Plant Mol. Biol. 11, 805-820 (1988)  
FEATURES  
source 1. .2346  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
join(196. .496,623. .888,985. .1431,1547. .1951)  
/note="12S storage protein CRA1"  
/codon\_start=1  
/protein\_id="AAA32777.1"  
/db\_xref="GI:166676"  
/translation="MARVSSLLSFCLTLFLFHGYAAQQGQGGQFFNECQLDQLNAL  
EPSHVLKSEAGRIEVDHHPALQRCGVSFARYIIIESKGLYLPSFNTAKLSFVAKGR  
GLMGKIPGCAETFDQSFQRFEGQSGQSRFDMHVKVEHIRSGDTIATTPGVAQM  
FYNDGQGPLVIVSFDLASHQNDLRNPRFYLAGNNPQGVWLGREQPQKNIFNG  
FGPEVTAQALKIDLTQAQQLQNDNRNIVRVGGPFGVIRPRLRGQPEEEEEER  
HGRNGLEETICSARCTNDLDDPSRADVYKPOLGYISTLSNYDLPLIRLIRLSALRG  
SIRQNAVLPMWNAANAILIETDGEAQIOIVNDNGNRVFDGVSGQLIAPQGFVS  
VKRATSNRFQWVEFTNNAQAINTLAGRTSVLRGLPLEVITNGFQISPEEARVKNT  
LETLTHSSGPASVGRPRVAA"  
exon <196. .496  
/note="12S storage protein CRA1"  
/number=1  
intron 497. .622  
/note="12S storage protein CRA1 intron A"  
exon 623. .888  
/number=2  
intron 889. .984  
/note="12S storage protein CRA1 intron B"  
exon 985. .1431  
intron 1432. .1546  
/note="12S storage protein CRA1 intron C"  
exon 1547. .>1951  
/note="12S storage protein CRA1"  
/number=4  
BASE COUNT 714 a 528 c 472 g 632 t  
ORIGIN  
Query Match 85.5%; Score 18.8; DB 8; Length 2346;  
Best Local Similarity 90.9%; Pred. No. 7.8;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 aacgagtgcagctagaccagc 22  
||||||| ||||| |||||  
Db 295 AACGAGTGCACCTGACCAGC 316  
RESULT 11  
LOCUS BNC2G 3081 bp DNA PLN 04-APR-1995  
DEFINITION B.napus Bnc2 gene for cruciferin storage protein.  
ACCESSION X59295  
VERSION X59295.1 GI:17791  
KEYWORDS cruciferin; cruciferin storage protein.  
SOURCE rape.  
ORGANISM Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 3081)  
AUTHORS Breen,J.P. and Crouch,M.L.  
TITLE Molecular analysis of a cruciferin storage protein gene family of Brassica napus  
JOURNAL Plant Mol. Biol. 19 (6), 1049-1055 (1992)  
MEDLINE 92379259  
REFERENCE 2 (bases 1 to 3081)  
AUTHORS Breen,J.P.  
TITLE Direct Submission

[illegible]



```
10646..10686,10782..10857))
/note="gene_id:MLN1.2"
/codon_start=1
/evidence="not_experimental"
/product="casein kinase I"
/protein_id="BAB10977.1"
/db_xref="GI:9759511"
/translation="MDLVIGGKFKLGGKIGSGFGELYGVNVTGEEVAVKLENVKT
KHPQLHYESKLYMLQGGSGIPNFKWFGEGDYVMVIDLLGSPLEDFNYCNKRKLTIL
KTVMLAQLLNLRVFMETGRFLHRDIKPDNFLMGLGRKANOVYIIDGLGKKYRDILQ
THKHTPYRENKMLTCTARYASVNTHLGVQSPRDDLESIGVYVLYELKSLPMWGLKA
GTTKKQYDRISEKKKYSTPIEVLCKNQPSSEFYSYFHYCSLRFDPDKPDSYXLRFLRDL
FTRGQYDIYVDVTVLKYPIQIGSSSGSSRTTHHTTAKPGFNADPIERQERILGET
TRYKIPGAVEAFSRHPHTTSPDRSRNSRDDGPFQKTHGDSEANSSRYRASS
SRKVAAGSSPSSAGPSTSSRLVSSSGGGGSGGNGRPS*SRVQAGYESKTL
SFSRATAARNTREDOLRSFELSLRK"
complement(join(12794..13054,13133..13219,13304..13440,
13520..13571,13656..13816,14094..14244))
/note="gb|AAD20643.1"
gene_id:MLN1.3"
/codon_start=1
/evidence="not_experimental"
/product="NED-like protein"
/protein_id="BAB10978.1"
/db_xref="GI:9759512"
/translation="MAEKNASVGDGAIIRVSGMQFSYDVQDPIFFDNLDLPAGSRCLL
VANGSGRTTLLIKLAGKHMVGKKNVQVLDRAFDHTELVCSGDL SylGGSWSKTAG
SAGDITPLOGDSABEHMIFGVEGIDPFRRKLDLDDINLQWRMHKVSVDGRRVVOICM
GLHHPFKVLLDVEDLDVVARMDLLEFFKEECRQRGATIVYAHIFDQRTWASHL
AYINGEKLKSLAKDEIKDLTSPNLLSVVEAWLKRSEKVEKTKHPVTSPPFMSSR
QMAYYR"
complement(join(15032..15436,15552..15998,16129..16394,
16518..16818))
/note="gene_id:MLN1.4"
/codon_start=1
/evidence="not_experimental"
/product="legumin-like protein"
/protein_id="BAB10979.1"
/db_xref="GI:9759513"
/translation="MARVSSLLSFCLTLLILPHGYAAQGGQGGQFPNECOLDQNAL
BPSHLKSEAGRIEVDHHAQPOLRCGVSFARYIIESGLXILPSPFNAKLSFVAKGR
GLMGKVIQCATFQDSSEFPQFEGQSGQSRFRMHQKVEHIRSGDIIATTPGVAQW
FYDGEQALVIVSDFDLASHQOLDNRNPPFYLAGNPGQVWLQGRQGPQKNIFFG
FPEYIAGALQITLQAOQLQONDNRNIVRVQPGFVIRPLRGQRPOEEEEEGR
HGRNGEETICSARCTNDLDPDRADYKPOLGYISTLNSYDLPILRFIRLSALRG
STROMAWLPQWNAANAILVYTDGEAQIQIVNDNGNRVFDGQVSGOLIAVPGQFSV
VKRATSNRQWVEFTKTNAQINTLAGRTSVLRGLPLEVITNGFQISPEEARVRENT
LETTLTHSGSPASYGRPRVAAA"
19700..20443
/note="contains similarity to surface protein"
gene_id:MLN1.5"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB10980.1"
/db_xref="GI:9759514"
/translation="WATPTLLLLTAVFLSTEITAAQRAAPGAPGAPINITAILEKG
GQFVTLRLNNTQIGNOINIQINSSSEGMTVLAPTDNAFQNLKPTLKNLSPDDQVK
LTLIYSPKFTLELLSVSNPRTQASGRGVYGLNFTGQGNQVNSVGTVETRL
STSLQRERPLAVYVDMVLLPEEMFGERKISPMAPPKSPDPVSDSSSKAAAPS
ESEKSGSEMNTGLGLIGLGVVLCIKFL"
join(21063..21237,21296..21408,21497..21667,21741..21953,
22037..22201)
/note="gene_id:MLN1.6"
/codon_start=1
/evidence="not_experimental"
/product="prohibitin"
/protein_id="BAB10981.1"
/db_xref="GI:9759515"
/translation="MNWKVPNVPGPSALSKLLGVITGGGLYICGSMYVNDGGRH
ALVNRFGDRLPEYITLQNGYGERVLIINETLKVAVQVNSQHLITQREAVREIR
KIVTERAKFNALDDVSTINLKFGKEFTEIERKQVAAQEAERAKFIVEAQDKKS
AIIIRAQGEAKSPALYICQAIANNEAFITLRKIEAAREIAQTIAKSANKVYLNSSDLLIS
```

KQ"

complement(join(22369..22845,22924..23036,23129..23230,23307..23673))

/note="gene\_id:MLN1.7

unknown protein"

/codon\_start=1

/evidence="not\_experimental

/protein\_id="BAB10982.1"

/db\_xref="GI:9759516"

/translation="MDSKSLAKSKRAHTLHSHKSHSVHKPKVPGVSEKNPEKLOGNQ

TKSPQSVRSVALPSNDRYDDELDAEDSSLSLHSDVIVPKSGADYVLIHLSEAQAE

SNKLENLNDLSSLDLHDEFVSVMISARGEGILSNMEDNDFVVEEDGSKVPG

FUSLNLADLTLENVDLHERLYIDPLLPPELNTSQTKVSRENEPSSHIAQNDPI

VPGFSEVREAESLDQKDIILITDESEKSAIEADLILLNSFEATQPPVPASAS

GKSSAFETELSLKSHSSTEQFNKPGNSDQKIHMTGFNDVLDLLESTPVSIIPOS

NOTSSKVLDDDFSDWLDTI"

join(31663..31801,32230..32626,33221..34085)

/note="contains similarity to zinc finger protein Id1

gene\_id:MLN1.8"

/codon\_start=1

/evidence="not\_experimental

/protein\_id="BAB10983.1"

/db\_xref="GI:9759517"

/translation="MTSEVLQTISSGSGFAQPSQSSITLDHDESLINPPLVKKRNLPG

NPDPAEVIALSPITLMATNRFCEVCGRQFQDNQLQHRHGNLPMKLDKORTSKEV

RKRVYCPKTCVHHHSRAGDLTGIRKHFCKRKEKWTCEKAKYAVGSDWKAH

SKTCGTRVRCDCGTIFSRRODSFTHRAFCDAEAETAKINAVSHLGAAGAPGSV

NINYOYLMGTFTIPLQFPVQPTNPNNHHOHFQPTSSLSLWNGQDIAPQPOPDY

DWFGNAKAASACIDNNYTHDQITQANASLTITTTLSAPSLFSSDQPPQANANSV

NNSATLQKAAEIGATSTTTAATNDPSTFLQSPFLKSDTQTTSDSGEKFALFGSN

NNIGLMSHSHDHOENARNNDVTYVSALDELQNYPWKRRRVDDGGGEGGGQTRDFLG

VGVQTLCHPSSINGWI"

38966..39670

/note="gene\_id:MLN1.9

pir||T38261

/codon\_start=1

/evidence="not\_experimental

/protein\_id="BAB10984.1"

/db\_xref="GI:9759518"

/translation="MKFTDSPVIDLTVMGTKLSIQDNGSMHVGTSVWPCSLILSKFA

ERNSTLDSSTTSNPFAELDFRRRGIELTGGCVAGMAFYLLGLTEIVLTIAP

VPMALKHLNRKNTALGSLKTSIVYNNRQDISALKPPFDLVIADVVYIIEESVGL

VPMALLVADGGAVLLGYQIRSPREADKLFWELCDLVFKIEKYPHEHLHSDYAEETDV

YIFRKKVKKNAEVS"

join(41744..41855,41947..42360,42441..42520,42621..43001,43096..43185,43267..43431,43518..43740,43847..43947,44084..44170,44251..44403,44498..44653,44728..44937,45026..45676,45764..46294,46387..46947,47129..47233,

Query Match 85.5%; Score 18.8; DB 8; Length 84544;

Best Local Similarity 90.9%; Pred. No. 9.2;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22

||||| ||||| |||||

Db 16719 AACGAGTGCCAGCTCGACCAGC 16698

RESULT 13

AC007864

LOCUS

AC007864 167645 bp DNA HTG 17-JUL-2001

DEFINITION

Trypanosoma brucei chromosome II clone RPI93-28H13, \*\*\* SEQUENCING

IN PROGRESS \*\*\*, 1 ordered pieces.

AC007864

AC007864.4 GI:14787195

VERSION

HTG; HTGS, PHASE2.

KEYWORDS

Trypanosoma brucei.

SOURCE

Trypanosoma brucei.

ORGANISM

Trypanosoma brucei

Eukaryota; Eulgenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE

1 (bases 1 to 167645)

AUTHORS

El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K.,



Db 918 GAGTCTCAGCTAGACCAG 935

Search completed: February 15, 2002, 18:55:19  
Job time: 20558 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:02:03 ; Search time 868.33 Seconds  
(without alignments)  
21.721 Million cell updates/sec

Title: US-09-698-903b-13

Perfect score: 22  
Sequence: 1 aacgagtgcagtagaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_ll01.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	22	100.0	22	AAH25457	PCR primer for end
2	22	100.0	22	AAH07002	PCR primer B03, to
3	22	100.0	3113	AAQ13870	Cruciferin A gene.
c 4	17.2	78.2	20633	AAH13213	Enterococcus faeca
c 5	16.8	76.4	1692	AAH60974	NLERK2 CDNA 3' reg
6	16.2	73.6	353	AAH15388	Pseudomonas aerugi
7	16.2	73.6	353	AAH15493	Pseudomonas aerugi
8	15.8	71.8	716	AAH05632	Human cDNA clone (
9	15.8	71.8	850	AAH07552	Human cDNA clone (
10	15.8	71.8	1790	AAH14119	Human cDNA sequenc
11	15.8	71.8	1929	AAH16554	Human cDNA sequenc

12	15.8	71.8	2096	22	AAH05091	Human secreted pro
13	15.8	71.8	10708	19	AAV69286	Sequence of mouse
c 14	15.6	70.9	163	22	AAI52779	Probe #21465 used
15	15.6	70.9	578	22	AAI39698	Probe #8384 used t
c 16	15.6	70.9	695	21	AAE12137	Aspergillus oryzae
c 17	15.6	70.9	1193	21	AAC41375	Arabidopsis thalia
c 18	15.6	70.9	1236	21	AAC40221	Arabidopsis thalia
c 19	15.6	70.9	1238	21	AAC49331	Arabidopsis thalia
c 20	15.6	70.9	1239	21	AAC49330	Arabidopsis thalia
c 21	15.6	70.9	1240	21	AAC33542	Arabidopsis thalia
c 22	15.6	70.9	1240	21	AAC33542	Arabidopsis thalia
c 23	15.6	70.9	1968	21	AAC6946	Human secreted pro
c 24	15.6	70.9	2105	22	AAI58985	Human polynucleoti
c 25	15.6	70.9	2224	22	AAI60771	Human polynucleoti
c 26	15.6	70.9	3192	21	AAC49328	Arabidopsis thalia
c 27	15.6	70.9	83390	21	AAF22283	BAC containing rep
c 28	15.6	70.9	90336	21	AAF22289	BAC containing rep
c 29	15.4	70.0	307	21	AAH43854	Human secreted exp
c 30	15.4	70.0	1256	21	AAZ60854	DNA encoding a kap
c 31	15.4	70.0	1330	16	AAQ75928	Mouse opioide recep
c 32	15.4	70.0	2518	21	AAZ60655	DNA encoding a kap
c 33	15.4	70.0	2600	16	AAQ90096	Mouse kappa-3 opio
c 34	15.4	70.0	2634	21	AAZ60653	DNA encoding a kap
c 35	15.2	69.1	4315	20	AAH12938	Enterococcus faeca
c 36	15.2	69.1	191	22	AAI55277	Probe #23963 used
c 37	15.2	69.1	350	13	AAQ31549	MAD-3. Homo sapie
c 38	15.2	69.1	520	22	AAI42192	Probe #10878 used
c 39	15.2	69.1	657	15	AAO54681	Rat proteosome RIN
c 40	15.2	69.1	754	22	AAH29516	Drosophila melanog
c 41	15.2	69.1	819	18	AAH91841	Human inhibitory k
c 42	15.2	69.1	930	20	AAH22769	Human SOCS14 cDNA.
c 43	15.2	69.1	1148	22	AAH26304	Pseudomonas sp Typ
c 44	15.2	69.1	1548	19	AAV55840	Human IkappaB poly
c 45	15.2	69.1	1550	13	AAQ31550	Encodes Ikb NF-kap
	15.2	69.1	1550	19	AAV41719	Human I-kappa-B-ai

ALIGNMENTS

RESULT 1  
AAH25457  
ID AAH25457 standard; DNA; 22 BP.  
XX  
AC AAH25457;  
XX  
XX  
DT 22-AUG-2001 (first entry)  
XX  
DE PCR primer for endogenous sequences in transgenic plants.  
XX  
KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
KW fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.  
XX  
OS Synthetic.  
XX  
XX WO200141558-A1.  
XX  
PD 14-JUN-2001.  
XX  
PF 06-DEC-2000; 2000WO-EF12872.  
XX  
PR 08-DEC-1999; 99US-0457037.  
XX  
PA (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
PI De Both G, De Beuckeleer M;  
XX  
DR WPI; 2001-381419/40.  
XX  
PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
PT with improved qualities, comprises a male-sterility gene and fertility  
PT restorer gene, integrated into the genome -  
XX  
PS Example 5; Page 53; 98pp; English.

XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAH25457-58 were used to amplify endogenous sequences  
 CC from transgenic plants of the invention.  
 XX Sequence 22 BP; 7 A; 6 C; 6 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 aacgagtgtcagctagaccagc 22  
 Db 1 aacgagtgtcagctagaccagc 22  
 |||||

RESULT 2  
 AAD07002  
 ID AAD07002 standard; DNA; 22 BP.  
 AC AAD07002;  
 XX  
 DT 06-AUG-2001 (first entry)  
 DE PCR primer B03, to recognise foreign DNA and flanking sequence of MS-B2.  
 KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; ss.  
 XX Brassica napus.

XX  
 PN W0200131042-A2.  
 XX  
 PD 03-MAY-2001.  
 XX  
 PF 26-OCT-2000; 2000WO-EP10680.  
 PR 29-OCT-1999; 99US-0430497.  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA Weston B, De Beuckeleer M;  
 PI WPI; 2001-300517/31.  
 DR Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -

XX Example 5; Page 33; 53pp; English.  
 PS The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.  
 XX Sequence 22 BP; 7 A; 6 C; 6 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 aacgagtgtcagctagaccagc 22  
 Db 1 aacgagtgtcagctagaccagc 22  
 |||||

RESULT 3  
 AAQ13870  
 ID AAQ13870 standard; DNA; 3113 BP.  
 AC AAQ13870;  
 XX  
 DT 09-DEC-1991 (first entry)  
 DE Cruciferin A gene.  
 XX Seed storage protein; crUA; ss.  
 KW Brassica napus.  
 OS  
 XX EP449376-A.  
 XX  
 PD 02-OCT-1991.  
 XX  
 PF 25-MAR-1991; 91EP-0200688.  
 PR 25-MAR-1991; 91EP-0200688.  
 PR 23-MAR-1990; 90US-0498561.  
 XX (KONN ) GIST-BROCADES NV.  
 PA Pen J, Sijmons PC, Van Ooyen AJJ, Rietveld K, Verwoerd TC;  
 PI Quax WJ;  
 XX WPI; 1991-289815/40.  
 DR Seeds contg. enhanced enzyme levels from transgenic plants - used  
 PT for catalysing reactions, increasing nutritional values or  
 PT treating digestive disorders.  
 XX Example; Fig 3; 38pp; English.  
 PS The DNA is the genomic sequence of the seed storage protein gene  
 CC cruciferin A (crUA). It can be used in the prodn. of transgenic  
 CC plants expressing cruciferin in its seeds for use in an industrial  
 CC process. The seeds contg. the cruciferin can be used without the  
 CC need for first extracting and/or isolating the enzymes. The use of  
 CC seeds for the storage of cruciferin provides a stable vehicle which  
 CC is easily packaged and transported and easily handled during use.  
 CC See also AAQ13871-Q13877.  
 XX  
 SQ Sequence 3113 BP; 961 A; 685 C; 586 G; 881 T; 0 other;

Query Match 100.0%; Score 22; DB 12; Length 3113;  
 Best Local Similarity 100.0%; Pred. No. 0.09;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 aacgagtgtcagctagaccagc 22  
 Db 790 aacgagtgtcagctagaccagc 811  
 |||||

RESULT 4  
 AAX13213/c  
 ID AAX13213 standard; DNA; 20633 BP.  
 AC AAX13213;  
 XX  
 DT 19-MAR-1999 (first entry)

```

XX DE Enterococcus faecalis genome contig SEQ ID NO:276.
XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;
XX KW vaccine; attenuation; computer readable medium; ds.
XX OS Enterococcus faecalis.
XX PN WO9850555-A2.
XX PD 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US08985.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Barash SC, Dillon PJ, Kunsch CA;
XX PI WPI; 1999-045171/04.
XX DR
XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides
XX PT - used to develop products for the detection of Enterococcus and for
XX PT use in vaccines for prevention or attenuation of Enterococcus
XX PT infection.
XX PS Claim 1; Page 1284-1294; 2084pp; English.
XX CC
XX CC A computer readable medium has been developed which has recorded on it
XX CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX CC AAX12938 to AAX13919 represent these nucleotide sequences which are
XX CC primary nucleotide sequences, also known as contigs. The computer-based
XX CC system can identify fragments of the Enterococcus faecalis genome with
XX CC commercial importance. The products can be used to detect the presence
XX CC of Enterococcus faecalis in samples. They can also be used for
XX CC diagnosing Enterococcal infection in an animal and monitoring
XX CC progression of disease, and for identifying agents which can be used to
XX CC modulate the growth or pathogenicity of Enterococcus faecalis, or
XX CC another related organism, in vivo or in vitro. In particular the
XX CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX CC can be used in vaccines to prevent or attenuate an Enterococcal
XX CC infection.
XX SQ Sequence 20633 BP; 5857 A; 4284 C; 3025 G; 7427 T; 40 other;

Query Match 78.2%; Score 17.2; DB 20; Length 20633;
Best Local Similarity 86.4%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||||| ||| ||| |||||
Db 18911 AACGAGTATCATCTAAACCAGC 18890

RESULT 5
AAT60974/C
ID AAT60974 standard; cDNA; 1692 BP.
XX AC
XX AC AAT60974;
XX DT 23-JUN-1997 (first entry)
XX DE
XX DE NLERK2 cDNA 3' region.
XX KW LERK; ligand for eph-related kinase; ERK; NLERK2;
XX KW receptor protein tyrosine kinase; cell proliferation;
XX KW cell differentiation; cell survival; nerve cell; ss.
XX OS Homo sapiens.

```

```

XX PN WO9704091-A1.
XX PD 06-FEB-1997.
XX PF 19-JUL-1996; 96WO-AU00460.
XX PR 05-FEB-1996; 96AU-0007890.
XX PR 20-JUL-1995; 95AU-0004263.
XX PR 27-NOV-1995; 95AU-0006847.
XX PR 22-DEC-1995; 95AU-0007299.
XX PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX PI Nicola NA;
XX PI WPI; 1997-132632/12.
XX PT Nucleic acid mol. encoding ligand for eph-related kinase - useful
XX PT for treatment of, pref. neuronal, cells to increase survival,
XX PT proliferation and differentiation
XX PS Claim 5; Page 43-45; 71pp; English.
XX CC
XX CC The 3' region (AAT60974) and 5' region (AAT60875) sequences were
XX CC detd. for a cDNA sequence (see also AAT60966) coding for NLERK2
XX CC (AAW10637), a novel human ligand for eph-related kinase (LERK). The
XX CC NLERK2 cDNA was obtd. from a human foetal brain cDNA library using
XX CC probes (see also AAT60967-69) based on an expressed sequence tag
XX CC previously isolated on the basis of homology to conserved regions
XX CC of known LERKs (see also AAW10633-36).
XX SQ Sequence 1692 BP; 316 A; 504 C; 385 G; 472 T; 15 other;

Query Match 76.4%; Score 16.8; DB 18; Length 1692;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acgagtgctcagctagaccag 21
||||||| ||| ||| |||||
Db 1434 ACCACTGTCAGCTAGACCAG 1415

RESULT 6
AAF15388
ID AAF15388 standard; DNA; 353 BP.
XX AC
XX AC AAF15388;
XX DT 09-MAR-2001 (first entry)
XX DE
XX DE Pseudomonas aeruginosa ssrA gene, SEQ ID NO:101.
XX KW ssrA gene; tmRNA; bacterium; chloroplast; diatom; detection;
XX KW detection; identification; quantification; characterisation;
XX KW nucleic acid array; DNA chip; drug design; treatment monitoring;
XX KW contamination; ds.
XX OS
XX OS Pseudomonas aeruginosa.
XX PN WO200070086-A1.
XX PD 23-NOV-2000.
XX PF 15-MAY-2000; 2000WO-IE000066.
XX PR 14-MAY-1999; 99WO-IE000043.
XX PA (IRBI-) ENTERPRISE IRELAND T/A BIORESEARCH IRELA.
XX PA (OYNA-) UNIV NAT IRELAND GALWAY.
XX PI Barry TG, Smith TJ;

```

XX WPI; 2001-025025/03.

XX Use of ssrA gene, tmRNA, or fragments of them, as target regions in

PT probe assays for detection of prokaryotic or eukaryotic organisms, and

PT for determination of species -

XX Disclosure; Page 55; 22lpp; English.

XX The invention relates to the ssrA gene or tmRNA, an RNA transcript

CC of the ssrA gene, or fragments thereof as target regions in a nucleic

CC acid probe assay for the detection and identification of prokaryotic

CC and/or eukaryotic organisms. The invention also relates to 38 novel ssrA

CC sequences and their tmRNA transcripts (claimed), and to pan-bacterial,

CC genus- and species-specific ssrA gene/tmRNA-directed probes and PCR

CC primers (claimed). tmRNA is a stable, high copy number RNA which is

CC found in all bacteria and is also found in chloroplasts and diatoms. It

CC has a dual function both as a tRNA and as an mRNA and is involved in

CC rescuing truncated mRNAs which have lost stop codons. SsrA genes and

CC tmRNA sequences can be used as target regions in nucleic acid probe

CC assays for the detection, identification, or quantification of a

CC prokaryotic or eukaryotic organism. cDNA transcripts of tmRNA molecules

CC may also be used as probes for in vitro or in situ nucleic acid

CC hybridisation assays. A fragment of the ssrA gene or a tmRNA molecule

CC corresponding to a region of high homology from the 5' end or the 3'

CC end of the DNA molecule can be used as a universal target region in

CC a nucleic acid probe assay, while a fragment of the ssrA gene or a

CC tmRNA molecule corresponding to a region of low homology can be used

CC as a target region to distinguish between species and as a target

CC region for the generation of genus-specific probes. These regions

CC may also be used as the basis for amplification primer design. The

CC target regions may be used as the basis of an assay for distinguishing

CC between living and dead prokaryotic or eukaryotic organisms, and in

CC a multiple probe format for broad scale detection and/or identification

CC of prokaryotic or eukaryotic organisms. An ssrA gene probe or a

CC tmRNA transcript probe can be linked to a microarray gene chip system

CC for the broad scale high throughput detection and identification of

CC prokaryotic or eukaryotic organisms. A fragment of the ssrA gene or

CC the tmRNA transcript can be used in an assay to obtain a DNA profile

CC of a prokaryotic organism and distinguish between strains of the same

CC species. The ssrA gene, the tmRNA transcript, DNA complementary to

CC an ssrA gene or tmRNA, or a fragment thereof can be used to design an

CC agent directed against infectious prokaryotic or eukaryotic organisms

CC for therapeutic purposes, and target regions may be used to monitor the

CC efficacy of drug therapies against infectious agents. Target regions may

CC also be used to monitor the viability and level of probiotic organisms

CC in the gastrointestinal tract. The methods and nucleic acids and

CC compositions of the invention have applications in medicine, and also

CC in industry (e.g., for assessing bacterial contamination of a foodstuff

CC or an environmental sample). Sequences AAF15338-F15442 represent ssrA

CC genes, or fragments thereof, from a wide variety of organisms.

XX Sequence 353 BP; 94 A; 94 C; 99 G; 66 T; 0 other;

XX

Query Match 73.6%; Score 16.2; DB 22; Length 353;

Best Local Similarity 85.7%; Pred. No. 57;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagacacg 21

Db 166 aacgagtgtcagctagacacg 186

||||| ||||| ||||| |||

RESULT 7

AAAF15493

ID AAF15493 standard; RNA; 353 BP.

XX

AC AAF15493;

XX

DT 09-MAR-2001 (first entry)

XX

DE Pseudomonas aeruginosa tmRNA, SEQ ID NO:102.

ssrA gene; tmRNA; bacterium; chloroplast; diatom; detection;

detection; identification; quantification; characterisation;

nucleic acid array; DNA chip; drug design; treatment monitoring;

contamination; ss.

Pseudomonas aeruginosa.

WO200070086-A1.

23-NOV-2000.

15-MAY-2000; 2000WO-IE000066.

14-MAY-1999; 99WO-IE000043.

(IRBI-) ENTERPRISE IRELAND T/A BIORESEARCH IRELA.

(UYNA-) UNIV NAT IRELAND GALWAY.

Barry TG, Smith TJ;

WPI; 2001-025025/03.

Use of ssrA gene, tmRNA, or fragments of them, as target regions in

probe assays for detection of prokaryotic or eukaryotic organisms, and

for determination of species -

Disclosure; Page 55; 22lpp; English.

The invention relates to the ssrA gene or tmRNA, an RNA transcript

of the ssrA gene, or fragments thereof as target regions in a nucleic

acid probe assay for the detection and identification of prokaryotic

and/or eukaryotic organisms. The invention also relates to 38 novel ssrA

sequences and their tmRNA transcripts (claimed), and to pan-bacterial,

genus- and species-specific ssrA gene/tmRNA-directed probes and PCR

primers (claimed). tmRNA is a stable, high copy number RNA which is

found in all bacteria and is also found in chloroplasts and diatoms. It

has a dual function both as a tRNA and as an mRNA and is involved in

rescuing truncated mRNAs which have lost stop codons. SsrA genes and

tmRNA sequences can be used as target regions in nucleic acid probe

assays for the detection, identification, or quantification of a

prokaryotic or eukaryotic organism. cDNA transcripts of tmRNA molecules

may also be used as probes for in vitro or in situ nucleic acid

hybridisation assays. A fragment of the ssrA gene or a tmRNA molecule

corresponding to a region of high homology from the 5' end or the 3'

end of the DNA molecule can be used as a universal target region in

a nucleic acid probe assay, while a fragment of the ssrA gene or a

tmRNA molecule corresponding to a region of low homology can be used

as a target region to distinguish between species and as a target

region for the generation of genus-specific probes. These regions

may also be used as the basis for amplification primer design. The

target regions may be used as the basis of an assay for distinguishing

between living and dead prokaryotic or eukaryotic organisms, and in

a multiple probe format for broad scale detection and/or identification

of prokaryotic or eukaryotic organisms. An ssrA gene probe or a

tmRNA transcript probe can be linked to a microarray gene chip system

for the broad scale high throughput detection and identification of

prokaryotic or eukaryotic organisms. A fragment of the ssrA gene or

the tmRNA transcript can be used in an assay to obtain a DNA profile

of a prokaryotic organism and distinguish between strains of the same

species. The ssrA gene, the tmRNA transcript, DNA complementary to

an ssrA gene or tmRNA, or a fragment thereof can be used to design an

agent directed against infectious prokaryotic or eukaryotic organisms

for therapeutic purposes, and target regions may be used to monitor the

efficacy of drug therapies against infectious agents. Target regions may

also be used to monitor the viability and level of probiotic organisms

in the gastrointestinal tract. The methods and nucleic acids and

compositions of the invention have applications in medicine, and also

in industry (e.g., for assessing bacterial contamination of a foodstuff

or an environmental sample). Sequences AAF15443-F15547 represent tmRNAs,

or fragments thereof, from a wide variety of organisms.

Sequence 353 BP; 94 A; 94 C; 99 G; 66 U; 0 other;



```

XX
SQ      Sequence 716 BP; 165 A; 196 C; 217 G; 131 T; 7 other;

      Query Match      71.8%; Score 15.8; DB 22; Length 716;
      Best Local Similarity 89.5%; Pred. No. 1e+02;
      Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 aacgagtcagctagaccagc 21
      ||||| :||| :||| |||
Db      166 aacgacugacagauagaacag 186

RESULT      8
AAH05632
ID      AAH05632 standard; cDNA; 716 BP.
XX
AC      AAH05632;
XX
DT      26-JUN-2001 (first entry)
XX
DE      Human cDNA clone (5'-primer) SEQ ID NO:2467.
XX
KW      Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS      Homo sapiens.
XX
PN      EP1074617-A2.
XX
PD      07-FEB-2001.
XX
PF      28-JUL-2000; 2000EP-0116126.
XX
PR      29-JUL-1999; 99JP-0248036.
XX
PR      27-AUG-1999; 99JP-0300253.
XX
PR      11-JAN-2000; 2000JP-0118776.
XX
PR      02-MAY-2000; 2000JP-0183767.
XX
PR      09-JUN-2000; 2000JP-0241899.
XX
PA      (HELI-) HELIX RES INST.
PI      Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI      Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR      WPI; 2001-318749/34.
XX
PT      Primer sets for synthesizing polynucleotides, particularly the 5602
PT      full-length cDNAs defined in the specification, and for the detection
PT      and/or diagnosis of the abnormality of the proteins encoded by the
XX      full-length cDNAs -
PS      Claim 1; SEQ ID 2467; 2537pp + CD ROM; English.
XX
CC      The present invention describes primer sets for synthesising 5602
CC      full-length cDNAs defined in the specification. Where a primer set
CC      comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC      to the complementary strand of a polynucleotide which comprises one of
CC      the 5602 nucleotide sequences defined in the specification, where the
CC      oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC      of an oligonucleotide comprising a sequence complementary to the
CC      complementary strand of a polynucleotide which comprises a 5'-end
CC      sequence and an oligonucleotide comprising a sequence complementary to a
CC      polynucleotide which comprises a 3'-end sequence, where the
CC      oligonucleotide comprises at least 15 nucleotides and the combination of
CC      the 5'-end sequence/3'-end sequence is selected from those defined in
CC      the specification. The primer sets can be used in antisense therapy and
CC      in gene therapy. The primers are useful for synthesising polynucleotides,
CC      particularly full-length cDNAs. The primers are also useful for the
CC      detection and/or diagnosis of the abnormality of the proteins encoded by
CC      the full-length cDNAs. The primers allow obtaining of the full-length
CC      cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC      AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC      AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC      represent oligonucleotides, all of which are used in the exemplification
CC      of the present invention.

XX
SQ      Sequence 716 BP; 165 A; 196 C; 217 G; 131 T; 7 other;

      Query Match      71.8%; Score 15.8; DB 22; Length 716;
      Best Local Similarity 89.5%; Pred. No. 1e+02;
      Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 gagtgctcagctagaccagc 22
      ||||| ||||| |||||
Db      122 gagtggcagctggaccagc 140

RESULT      9
AAH07552
ID      AAH07552 standard; cDNA; 850 BP.
XX
AC      AAH07552;
XX
DT      26-JUN-2001 (first entry)
XX
DE      Human cDNA clone (5'-primer) SEQ ID NO:4387.
XX
KW      Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS      Homo sapiens.
XX
PN      EP1074617-A2.
XX
PD      07-FEB-2001.
XX
PF      28-JUL-2000; 2000EP-0116126.
XX
PR      29-JUL-1999; 99JP-0248036.
XX
PR      27-AUG-1999; 99JP-0300253.
XX
PR      11-JAN-2000; 2000JP-0118776.
XX
PR      02-MAY-2000; 2000JP-0183767.
XX
PR      09-JUN-2000; 2000JP-0241899.
XX
PA      (HELI-) HELIX RES INST.
PI      Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI      Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR      WPI; 2001-318749/34.
XX
PT      Primer sets for synthesizing polynucleotides, particularly the 5602
PT      full-length cDNAs defined in the specification, and for the detection
PT      and/or diagnosis of the abnormality of the proteins encoded by the
XX      full-length cDNAs -
PS      Claim 1; SEQ ID 4387; 2537pp + CD ROM; English.
XX
CC      The present invention describes primer sets for synthesising 5602
CC      full-length cDNAs defined in the specification. Where a primer set
CC      comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC      to the complementary strand of a polynucleotide which comprises one of
CC      the 5602 nucleotide sequences defined in the specification, where the
CC      oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC      of an oligonucleotide comprising a sequence complementary to the
CC      complementary strand of a polynucleotide which comprises a 5'-end
CC      sequence and an oligonucleotide comprising a sequence complementary to a
CC      polynucleotide which comprises a 3'-end sequence, where the
CC      oligonucleotide comprises at least 15 nucleotides and the combination of
CC      the 5'-end sequence/3'-end sequence is selected from those defined in
CC      the specification. The primer sets can be used in antisense therapy and
CC      in gene therapy. The primers are useful for synthesising polynucleotides,
CC      particularly full-length cDNAs. The primers are also useful for the
CC      detection and/or diagnosis of the abnormality of the proteins encoded by
CC      the full-length cDNAs. The primers allow obtaining of the full-length
CC      cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC      AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC      AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC      represent oligonucleotides, all of which are used in the exemplification
CC      of the present invention.
```

CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 850 BP; 196 A; 227 C; 258 G; 160 T; 9 other;

Query Match 71.8%; Score 15.8; DB 22; Length 850;  
 Best Local Similarity 89.5%; Pred. No. 1e+02; Indels 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 2;  
 QY 4 gagtgctagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 126 gagtggcagctggaccagc 144

RESULT 10  
 AAH14119  
 ID AAH14119 standard; cDNA; 1790 BP.  
 XX  
 AC AAH14119;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:11310.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 XX

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 11310; 2537pp + CD ROM; English.  
 The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the polynucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 1790 BP; 398 A; 456 C; 504 G; 432 T; 0 other;

Query Match 71.8%; Score 15.8; DB 22; Length 1790;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gagtgctagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 122 gagtggcagctggaccagc 140

RESULT 11  
 AAH16554  
 ID AAH16554 standard; cDNA; 1929 BP.  
 XX  
 AC AAH16554;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:15615.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 XX

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 15615; 2537pp + CD ROM; English.  
 The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the polynucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by



FT intron 1463..1994  
 FT /\*tag= c  
 FT /number= 1  
 FT /note= "activin beta c partial intron (12-kb)"  
 FT 1995..2733  
 FT /\*tag= d  
 FT /number= 2  
 FT 2737..8248  
 FT /\*tag= e  
 FT /note= "activin beta e promoter region"  
 FT 8249..9536  
 FT /\*tag= f  
 FT /product= "activin beta e"  
 FT /note= "contains introns"  
 FT 8249..8546  
 FT /\*tag= g  
 FT /number= 1  
 FT /note= "activin beta e exon 1"  
 FT 8547..8782  
 FT /\*tag= h  
 FT /number= 1  
 FT /note= "activin beta e intron"  
 FT 8783..9533  
 FT /\*tag= i  
 FT /number= 2  
 FT /note= "activin beta e exon 2"  
 PN W09822492-A1.  
 XX 28-MAY-1998.  
 XX 20-NOV-1997; 97WO-US20882.  
 XX 20-NOV-1996; 96US-0752919.  
 XX (UNMI ) UNIV MICHIGAN.  
 XX Bonadio J, Fang J;  
 XX WPI; 1998-312408/27.  
 XX P-PSDB; AAW60617, AAW60618.  
 XX  
 XX New isolated nucleic acid encoding sub-units of liver activin -  
 XX useful for regulating growth and differentiation of cells, e.g. for  
 XX treating liver, bone and haematopoietic disorders  
 XX  
 XX Disclosure; Fig 4D-G; 141pp; English.  
 XX  
 XX This represents the sequence of mouse activin genetic loci. The invention  
 XX relates to murine beta c and beta e polypeptides and the genes encoding  
 XX them. Disorders of cell growth or differentiation (or susceptibility to  
 XX them) are diagnosed by measuring liver activin gene activity or by  
 XX detecting a mutation in the liver activin gene. Disorders of  
 XX haematopoiesis, erythroid differentiation, ovarian follicular maturation,  
 XX hormone secretion, neuronal survival, spermatogenesis, bone formation,  
 XX insulin secretion or cardiac morphogenesis are some conditions that can  
 XX be diagnosed using the liver activin. Cell growth and differentiation can  
 XX be stimulated by treatment with an liver activin compound or agent that  
 XX upregulates the compound's expression. Antagonists can be used to treat  
 XX liver diseases while agonists can be used to increase growth and  
 XX regeneration of liver tissue. The liver activin compound may also induce  
 XX bone growth (e.g. for treating osteoporosis or osteomalacia) or  
 XX haematopoiesis, particularly erythropoiesis, e.g. for treating  
 XX haemophilia, cystic fibrosis or menstrual disorders. Antibodies are  
 XX useful in immunoassays, to generate anti-idiotypic antibodies (which bind  
 XX to liver activin receptors), and to inhibit liver activin. Also,  
 XX transgenic animals containing liver activin gene can be used to produce  
 XX the liver activin (in correctly processed and modified forms) proteins,  
 XX or the transgenic animals, are useful for screening for liver activin  
 XX modulators.  
 XX This cDNA encodes a murine liver activin beta c polypeptide. Sequences  
 XX derived from beta c cDNA clone is used for screening and cloning a liver  
 XX activin beta e gene. Disorders of cell growth or differentiation (or

CC susceptibility to them) are diagnosed by measuring liver activin gene  
 CC activity or by detecting a mutation in the liver activin gene. Disorders  
 CC of haematopoiesis, erythroid differentiation, ovarian follicular  
 CC maturation, hormone secretion, neuronal survival, spermatogenesis, bone  
 CC formation, insulin secretion or cardiac morphogenesis are some conditions  
 CC that can be diagnosed using the liver activin. Cell growth and  
 CC differentiation can be stimulated by treatment with an liver activin  
 CC compound or agent that upregulates the compound's expression. Antagonists  
 CC can be used to treat liver diseases while agonists can be used to  
 CC increase growth and regeneration of liver tissue. The liver activin  
 CC compound may also induce bone growth (e.g. for treating osteoporosis or  
 CC osteomalacia) or haematopoiesis, particularly erythropoiesis, e.g. for  
 CC treating haemophilia, cystic fibrosis or menstrual disorders. Antibodies  
 CC are useful in immunoassays, to generate anti-idiotypic antibodies (which  
 CC bind to liver activin receptors) and to inhibit liver activin. Also,  
 CC transgenic animals containing liver activin gene can be used to produce  
 CC the liver activin (in correctly processed and modified forms) proteins,  
 CC or the transgenic animals, are useful for screening for liver activin  
 CC modulators.

SQ Sequence 10708 BP; 2651 A; 2844 C; 2607 G; 2546 T; 60 other;

Query Match 71.8%; Score 15.8; DB 19; Length 10708;  
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gaggtgcagtagaccagc 22  
 ||||| ||||| |||||

Db 1949 GAGTGTCTGCTGCAGCAGC 1931

RESULT 14

AAI52779

ID AAI52779 standard; DNA; 163 BP.

AC AAI52779;

XX 17-OCT-2001 (first entry)

DE Probe #21465 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

XX Homo sapiens.

OS WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488997/53.

DR Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 21465; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SEN).

CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.

Search completed: February 15, 2002, 19:02:07  
Job time: 20771 sec

XX Sequence 163 BP; 30 A; 44 C; 50 G; 39 T; 0 other;

Query Match 70.9%; Score 15.6; DB 22; Length 163;  
Best Local Similarity 81.8%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22  
|| ||| || ||||| |||||  
Db 130 aaggagttctgctagaccagc 151

## RESULT 15

AAI39698  
ID AAI39698 standard; DNA; 578 BP.

XX AC AAI39698;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #8384 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;  
genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2..

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -

XX PS Claim 25; SEQ ID NO 8384; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP).

XX CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 578 BP; 145 A; 139 C; 145 G; 149 T; 0 other;

Query Match 70.9%; Score 15.6; DB 22; Length 578;  
Best Local Similarity 81.8%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22  
|| ||| || ||||| |||||  
Db 257 aaggagttctgctagaccagc 278

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:09:18 ; Search time 353.79 Seconds  
(without alignments)  
14.083 Million cell updates/sec

Title: US-09-698-903b-13

Perfect score: 22

Sequence: 1 aacgagtgcagtagaccagc 22

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	3113	1	US-08-146-422-20
2	22	100.0	3113	1	US-08-626-554-2
3	15.4	70.0	1330	3	US-08-147-592A-5
4	15.4	70.0	2600	1	US-08-147-949A-1
5	15.2	69.1	350	2	US-08-466-337A-16
6	15.2	69.1	350	2	US-08-475-359-16
7	15.2	69.1	350	3	US-08-465-887A-16
8	15.2	69.1	819	4	US-09-349-627-2
9	15.2	69.1	1550	2	US-08-466-337A-17
10	15.2	69.1	1550	2	US-08-475-359-17
11	15.2	69.1	1550	2	US-08-802-322-2
12	15.2	69.1	1550	3	US-08-465-887A-17
13	15.2	69.1	1550	3	US-08-895-601-3
14	14.6	66.4	2028	2	US-08-933-750C-73
15	14.6	66.4	2028	3	US-09-234-613-73
16	14.6	66.4	2915	2	US-08-500-857A-1
17	14.6	66.4	3719	1	US-08-920-812-10
18	14.6	66.4	3719	1	US-08-920-827-10
19	14.6	66.4	3719	1	US-08-921-177-10
20	14.6	66.4	3719	1	US-08-362-577C-10
21	14.6	66.4	3719	2	US-08-920-828-10
22	14.6	66.4	50341	1	US-08-247-901C-1
23	14.6	66.4	50341	2	US-09-075-904-1
24	14.6	66.4	52297	4	US-09-426-436-1
25	14.6	66.4	52297	4	US-08-705-557-1
26	14.6	66.4	4411529	4	US-09-103-840A-1
27	14.4	65.5	1458	5	PCT-US94-03437-1

c	28	14.4	65.5	1461	1	US-07-596-867C-10	Sequence 10, Appl
c	29	14.4	65.5	1506	3	US-08-445-463B-1	Sequence 1, Appli
c	30	14.4	65.5	1506	3	US-08-445-464C-1	Sequence 1, Appli
c	31	14.4	65.5	1963	5	PCT-US91-07715A-1	Sequence 1, Appli
c	32	14.4	65.5	1970	1	US-07-596-867C-1	Sequence 1, Appli
c	33	14.2	64.5	192	4	US-09-060-756-634	Sequence 634, App
c	34	14.2	64.5	215	4	US-09-060-756-194	Sequence 194, App
c	35	14.2	64.5	230	4	US-09-060-756-422	Sequence 422, App
c	36	14.2	64.5	579	2	US-08-611-757-50	Sequence 50, Appl
c	37	14.2	64.5	579	5	PCT-US95-05980-50	Sequence 50, Appl
c	38	14.2	64.5	1679	6	5196194-12	Patent No. 5196194
c	39	14.2	64.5	2389	2	US-08-372-652-8	Sequence 8, Appli
c	40	14.2	64.5	2389	5	PCT-US95-16311-8	Sequence 8, Appli
c	41	14.2	64.5	2940	3	US-09-195-868-11	Sequence 11, Appl
c	42	14.2	64.5	3262	2	US-08-678-039A-41	Sequence 41, Appl
c	43	14.2	64.5	3691	3	US-09-195-868-12	Sequence 12, Appl
c	44	14.2	64.5	4147	3	US-08-560-005-1	Sequence 1, Appli
c	45	14.2	64.5	4147	4	US-09-418-540-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-146-422-20  
; Sequence 20, Application US/08146422  
; Patent No. 5543576  
; GENERAL INFORMATION:  
; APPLICANT: VAN OOLJEN, ALBERT J. J.  
; APPLICANT: RIETVELD, KRIJN  
; APPLICANT: HOEKEMA, ANDREAS  
; APPLICANT: PEN, JAN  
; APPLICANT: SIJMONS, PETER C.  
; APPLICANT: VERWOERD, TEUNIS C.  
; APPLICANT: QUAX, WILHEMUS J.  
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
; TITLE OF INVENTION: USE  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/146,422  
; FILING DATE: 02-NOV-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KENNEDY, BILL  
; REGISTRATION NUMBER: 33,407  
; REFERENCE/DOCKET NUMBER: 44615-20011.23  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: nucleic acid  
; LENGTH: 3113 base pairs  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-146-422-20  
  
Query Match 100.0%; Score 22; DB 1; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.042;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgctagctagaccagc 22  
 |||||  
 Db 790 AACGAGTGTCTAGCTAGACCAGC 811

## RESULT 2

US-08-626-554-2  
 ; Sequence 2, Application US/08626554  
 ; Patent No. 5714474  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN COIJEN, ALBERT J.J.  
 ; APPLICANT: RIETVELD, ANDREAS  
 ; APPLICANT: HOEKEMA, KRIJN  
 ; APPLICANT: PEN, JAN  
 ; APPLICANT: SIJMONS, PETER C.  
 ; APPLICANT: VERWOERD, TEUNIS C.  
 ; APPLICANT: QUAX, WILHEMUS J.  
 ; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
 ; TITLE OF INVENTION: USE  
 ; NUMBER OF SEQUENCES: 32  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 PENNSYLVANIA AVENUE NW  
 ; CITY: WASHINGTON  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1888  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/626,554  
 ; FILING DATE: 02-APR-1996  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 26192-20011.10  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030 MRSNFOERSWSH  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3113 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-626-554-2

Query Match 100.0%; Score 22; DB 1; Length 3113;  
 Best Local Similarity 100.0%; Pred. No. 0.042;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgctagctagaccagc 22  
 |||||  
 Db 790 AACGAGTGTCTAGCTAGACCAGC 811

## RESULT 3

US-08-147-592A-5/C  
 ; Sequence 5, Application US/08147592A  
 ; Patent No. 6096513  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bell, Graeme I  
 ; APPLICANT: Reisine, Terry  
 ; APPLICANT: Yasuda, Kazuki  
 ; TITLE OF INVENTION: Opioid Receptor Genes,

; TITLE OF INVENTION: Compositions and Methods  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold, White & Durkee  
 ; STREET: P.O. Box 4433  
 ; CITY: Houston  
 ; STATE: Texas  
 ; COUNTRY: United States of America  
 ; ZIP: 77210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/147,592A  
 ; FILING DATE: 05-NOV-1993  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/100,694  
 ; FILING DATE: 30-JUL-1993  
 ; CLASSIFICATION: 435  
 ; APPLICATION NUMBER: 08/066,296  
 ; FILING DATE: 20-MAY-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Wilson, Mark B.  
 ; REGISTRATION NUMBER: 37,259  
 ; REFERENCE/DOCKET NUMBER: ARCD:105  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (512) 418-3000  
 ; TELEFAX: (512) 474-7577  
 ; TELEX: N/A  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1330 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 161..1261  
 ; US-08-147-592A-5

Query Match 70.0%; Score 15.4; DB 3; Length 1330;  
 Best Local Similarity 94.1%; Pred. No. 63;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 agtgtcagctagaccag 21  
 |||||  
 Db 462 AGTGTCTAGCTAGACCAG 446

## RESULT 4

US-08-147-949A-1/c  
 ; Sequence 1, Application US/08147949A  
 ; Patent No. 5747279  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pasternak, Gavril W.  
 ; APPLICANT: Pan, Ying-Xian  
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
 ; TITLE OF INVENTION: KAPPA3 OPIOID RECEPTORS RECEPTORS  
 ; TITLE OF INVENTION: ENCODED THEREBY, AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036



```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,949A
; FILING DATE: 05-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 44782/JPW/JKM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 299..1401
; OTHER INFORMATION:
; US-08-147-949A-1

Query Match 70.0%; Score 15.4; DB 1; Length 2600;
Best Local Similarity 94.1%; Pred. No. 66;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 agtgtcagctagaccag 21
| | | | | | | | | | | | | | | |
Db 600 AGTGTACGACGACCAG 584

RESULT 5
US-08-466-337A-16/c
; Sequence 16, Application US/08466337A
; Patent No. 5830756
; GENERAL INFORMATION:
; APPLICANT: Haskill, John S.
; APPLICANT: Baldwin Jr., Albert S.
; APPLICANT: Ralph, Peter
; TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/ 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,359
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 0899.004.33514
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEX: 312/474-0448
; TELEFAX: 25-3856
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-475-359-16

Query Match 69.1%; Score 15.2; DB 2; Length 350;
Best Local Similarity 85.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgtcagctagaccagc 22
| | | | | | | | | | | | | | | |
Db 135 CTAGTGTGAGCTGCCGACG 116

RESULT 6
US-08-475-359-16/c
; Sequence 16, Application US/08475359
; Patent No. 5846714
; GENERAL INFORMATION:
; APPLICANT: Haskill, John S.
; APPLICANT: Baldwin Jr., Albert S.
; APPLICANT: Ralph, Peter
; TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/ 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,359
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 0899.004.33514
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEX: 312/474-0448
; TELEFAX: 25-3856
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-475-359-16

Query Match 69.1%; Score 15.2; DB 2; Length 350;
Best Local Similarity 85.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgtcagctagaccagc 22
| | | | | | | | | | | | | | | |
```

```
Db 135 CTAGTGTGCTGCTGGCCGAC 116
RESULT 7
US-08-465-887A-16/c
; Sequence 16, Application US/08465887A
; Patent No. 6001582
; GENERAL INFORMATION:
; APPLICANT: Haskill, John S.
; APPLICANT: Baldwin Jr., Albert S.
; APPLICANT: Ralph, Peter
; TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional
; TITLE OF INVENTION: Activator and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/ 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,887A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 0899.006/33516
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-465-887A-16

Query Match 69.1%; Score 15.2; DB 3; Length 350;
Best Local Similarity 85.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgctcagctagaccagc 22
Db 135 CTAGTGTGCTGCTGGCCGAC 116
RESULT 8
US-09-627-2/c
; Sequence 2, Application US/09349627
; Patent No. 6194175
; GENERAL INFORMATION:
; APPLICANT: Peltz, Gary A.
; APPLICANT: Belloni, Paula N.
; TITLE OF INVENTION: Truncated Forms of Inhibitory kappa B
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Roche Bioscience
; STREET: 3401 Hillview Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

Query Match 69.1%; Score 15.2; DB 4; Length 819;
Best Local Similarity 85.0%; Pred. No. 76;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgctcagctagaccagc 22
Db 691 CTAGTGTGCTGCTGGCCGAC 672
RESULT 9
US-08-466-337A-17/c
; Sequence 17, Application US/08466337A
; Patent No. 5830756
; GENERAL INFORMATION:
; APPLICANT: Haskill, John S.
; APPLICANT: Baldwin Jr., Albert S.
; APPLICANT: Ralph, Peter
; TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional
; TITLE OF INVENTION: Activator and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/ 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,337A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 0899.008/33518
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 95..1045  
US-08-466-337A-17

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgtagcagcaccg 22  
| ||||| |||||  
Db 917 CTAGTGTGAGTGGCCGAGC 898

## RESULT 10

US-08-475-359-17/c  
Sequence 17, Application US/08475359  
Patent No. 5846714  
GENERAL INFORMATION:  
APPLICANT: Haskill, John S.  
APPLICANT: Baldwin Jr., Albert S.  
APPLICANT: Ralph, Peter  
TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional  
TITLE OF INVENTION: Activator and Uses Thereof  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower/ 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,359  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pochoplen, Donald J.  
REGISTRATION NUMBER: 32,167  
REFERENCE/DOCKET NUMBER: 0899,004.33514  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 95..1045  
US-08-475-359-17

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgtagcagcaccg 22  
| ||||| |||||  
Db 917 CTAGTGTGAGTGGCCGAGC 898

## RESULT 11

US-08-802-322-2/c  
Sequence 2, Application US/08802322  
Patent No. 5932425  
GENERAL INFORMATION:  
APPLICANT: Alkalay, Irit  
APPLICANT: Ben-Neriah, Yaron  
APPLICANT: Ciechanover, Aaron  
APPLICANT: Manning, Anthony  
APPLICANT: Mercurio, Frank  
APPLICANT: Yaron, Avraham  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING  
TITLE OF INVENTION: CELLULAR NF-(B ACTIVATION  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,322  
FILING DATE: 18-FEB-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 860098.418  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-802-322-2

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgtagcagcaccg 22  
| ||||| |||||  
Db 917 CTAGTGTGAGTGGCCGAGC 898

## RESULT 12

US-08-465-887A-17/c  
Sequence 17, Application US/08465887A  
Patent No. 6001582  
GENERAL INFORMATION:  
APPLICANT: Haskill, John S.  
APPLICANT: Baldwin Jr., Albert S.  
APPLICANT: Ralph, Peter  
TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional.

```
; TITLE OF INVENTION: Activator and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/ 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,887A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 0899.006/33516
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEX: 312/474-0448
; FAX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1550 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..1045
; US-08-465-887A-17

Query Match 69.1%; Score 15.2; DB 3; Length 1550;
Best Local Similarity 85.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgtagctagaccagc 22
| | | | | | | | | | | | | | | |
Db 917 CTAGTGTCTAGCTGGCCAGC 898

RESULT 13
US-08-995-601-3/c
; Sequence 3, Application US/08895601
; Patent No. 6060262
; GENERAL INFORMATION:
; APPLICANT: Beer-Romero, Peggy
; APPLICANT: Strack, Peter J.
; APPLICANT: Glass, Susan J.
; APPLICANT: Rolfe, Mark
; TITLE OF INVENTION: REGULATION OF KAPPA B (IKB) DEGRADATION,
; AND METHODS AND REAGENTS RELATED THERETO
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/895,601
; FILING DATE: 16-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MIV-096.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1550 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..1045
; US-08-895-601-3

Query Match 69.1%; Score 15.2; DB 3; Length 1550;
Best Local Similarity 85.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cgagtgtagctagaccagc 22
| | | | | | | | | | | | | | | |
Db 917 CTAGTGTCTAGCTGGCCAGC 898

RESULT 14
US-08-933-750C-73/C
; Sequence 73, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:04:21 ; Search time 9904.61 Seconds  
(without alignments)  
23.868 Million cell updates/sec

Title: US-09-698-903B-13

Perfect score: 22

Sequence: 1 aacgagtgcagtagaccagg 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estfun:\*\*
  - 2: em\_esthum:\*\*
  - 3: em\_estin:\*\*
  - 4: em\_estom:\*\*
  - 5: em\_estpl:\*\*
  - 6: em\_estba:\*\*
  - 7: em\_estro:\*\*
  - 8: em\_estov:\*\*
  - 9: em\_hc:\*\*
  - 10: gb\_est1:\*\*
  - 11: gb\_est2:\*\*
  - 12: gb\_hc:\*\*
  - 13: gb\_gss:\*\*
  - 14: em\_gss\_fun:\*\*
  - 15: em\_gss\_hum:\*\*
  - 16: em\_gss\_inv:\*\*
  - 17: em\_gss\_pln:\*\*
  - 18: em\_gss\_pro:\*\*
  - 19: em\_gss\_rod:\*\*
  - 20: em\_gss\_vrt:\*\*
  - 21: em\_gss\_other:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.8	85.5	152	10	BE520593
2	18.8	85.5	174	10	BE523925
3	18.8	85.5	182	10	BE524190
4	18.8	85.5	206	10	BE525280
5	18.8	85.5	212	10	BE523695
6	18.8	85.5	223	10	BE524251
7	18.8	85.5	225	10	BE523875
8	18.8	85.5	231	10	BE522306
9	18.8	85.5	232	10	BE520764
10	18.8	85.5	232	10	BE525569
11	18.8	85.5	238	10	BE525596
12	18.8	85.5	239	10	BE523203

13	18.8	85.5	243	10	BE525564
14	18.8	85.5	244	10	BE525454
15	18.8	85.5	244	10	BE525533
16	18.8	85.5	245	10	BE525453
17	18.8	85.5	247	10	BE523445
18	18.8	85.5	249	10	BE525542
19	18.8	85.5	252	10	BE524837
20	18.8	85.5	253	10	BE525432
21	18.8	85.5	253	10	BE525500
22	18.8	85.5	254	10	BE522200
23	18.8	85.5	254	10	BE524936
24	18.8	85.5	255	10	BE525465
25	18.8	85.5	261	10	BE522627
26	18.8	85.5	272	10	BE523398
27	18.8	85.5	277	10	BE524178
28	18.8	85.5	280	10	BE520917
29	18.8	85.5	283	10	BE520827
30	18.8	85.5	287	10	BE522480
31	18.8	85.5	291	10	BE521713
32	18.8	85.5	298	10	BE523183
33	18.8	85.5	300	10	BE525377
34	18.8	85.5	303	10	BE525401
35	18.8	85.5	305	10	BE520824
36	18.8	85.5	305	10	BE521646
37	18.8	85.5	305	10	BE524379
38	18.8	85.5	306	10	BE522847
39	18.8	85.5	306	10	BE523254
40	18.8	85.5	307	10	BE523644
41	18.8	85.5	308	10	BE524918
42	18.8	85.5	310	10	BE520427
43	18.8	85.5	311	10	BE520918
44	18.8	85.5	312	10	BE520556
45	18.8	85.5	312	10	BE522759

ALIGNMENTS

RESULT	1
BE520593	
LOCUS	BE520593 152 bp mRNA EST 19-MAR-2001
DEFINITION	M13F7/STM Arabidopsis developing seed Arabidopsis thaliana cDNA clone M13F7 5', mRNA sequence.
ACCESSION	BE520593
VERSION	BE520593.1 GI:9778571
KEYWORDS	EST.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	1 (bases 1 to 152)
AUTHORS	White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE	A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil plant physiol. 124 (4), 1582-1594 (2000)
JOURNAL	20567808
MEDLINE	Contact: Benning, C
COMMENT	Dept. of Biochemistry & Molecular Biology Michigan State University 224 Biochemistry, Michigan State University, East Lansing, MI 48824 , USA Tel: 517 355 1609 Fax: 517 353 9334 Email: benning@msu.edu Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371. Location/Qualifiers 1. .152 /organism="Arabidopsis thaliana"
FEATURES	
source	

```

/strain="Columbia"
/db_xref="taxon:3702"
/clone="M13F7"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
34 t

BASE COUNT      29 a 47 c 42 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 152;
Best Local Similarity 90.9%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 2
BE523925      174 bp mRNA EST 19-MAR-2001
LOCUS M43F11STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M43F11 5', mRNA sequence.
ACCESSION BE523925
VERSION BE523925
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .174
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F11"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
46 t

BASE COUNT      36 a 52 c 40 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 174;
Best Local Similarity 90.9%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 2
BE523925      174 bp mRNA EST 19-MAR-2001
LOCUS M43F11STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M43F11 5', mRNA sequence.
ACCESSION BE523925
VERSION BE523925
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .174
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F11"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
46 t

BASE COUNT      36 a 52 c 40 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 174;
Best Local Similarity 90.9%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 2
BE523925      174 bp mRNA EST 19-MAR-2001
LOCUS M43F11STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M43F11 5', mRNA sequence.
ACCESSION BE523925
VERSION BE523925
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .174
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F11"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
46 t

BASE COUNT      36 a 52 c 40 g
ORIGIN

```

```

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 104 AACGAGTCCAGCTCGACCAGC 125

RESULT 3
BE524190      182 bp mRNA EST 19-MAR-2001
LOCUS M47A2STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M47A2 5', mRNA sequence.
ACCESSION BE524190
VERSION BE524190
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .182
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M47A2"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
42 t

BASE COUNT      33 a 58 c 49 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 182;
Best Local Similarity 90.9%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 4
BE525280      206 bp mRNA EST 19-MAR-2001
LOCUS M61A10STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M61A10 5', mRNA sequence.
ACCESSION BE525280
VERSION BE525280
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .182
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M47A2"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
42 t

BASE COUNT      33 a 58 c 49 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 182;
Best Local Similarity 90.9%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 4
BE525280      206 bp mRNA EST 19-MAR-2001
LOCUS M61A10STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M61A10 5', mRNA sequence.
ACCESSION BE525280
VERSION BE525280
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .182
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M47A2"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
42 t

BASE COUNT      33 a 58 c 49 g
ORIGIN

```



Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 206)  
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
 A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
 Plant Physiol. 124 (4), 1582-1594 (2000)  
 20567808  
 Contact: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 , USA  
 Tel: 517 355 1609  
 Fax: 517 353 9334  
 Email: benning@msu.edu  
 Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

# FEATURES

source  
 1. .206  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="M61A10"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"  
 /note="Organ: Developing seed; Vector: pBluescript SK-;  
 Site\_1: EcoRI; Site\_2: XhoII"  
 BASE COUNT 44 a 58 c 54 g 50 t  
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 206;  
 Best Local Similarity 90.9%; Pred. No. 82;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22  
 ||||| |||| |||||  
 Db 121 AACGAGTGCACGCTCGACCAGC 142

# RESULT 5

BE523695 212 bp mRNA EST 19-MAR-2001  
 LOCUS M40E7STM Arabidopsis developing seed Arabidopsis thaliana CDNA  
 DEFINITION clone M40E7 5', mRNA sequence.  
 ACCESSION BE523695  
 VERSION BE523695.1 GI:9781673  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 212)  
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
 A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
 Plant Physiol. 124 (4), 1582-1594 (2000)  
 20567808  
 Contact: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 , USA  
 Tel: 517 355 1609

# REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

Fax: 517 353 9334  
 Email: benning@msu.edu  
 Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

# FEATURES

source  
 1. .212  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="M40E7"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"  
 /note="Organ: Developing seed; Vector: pBluescript SK-;  
 Site\_1: EcoRI; Site\_2: XhoII"  
 BASE COUNT 42 a 73 c 45 g 52 t  
 ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 212;  
 Best Local Similarity 90.9%; Pred. No. 83;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22  
 ||||| |||| |||||  
 Db 24 AACGAGTGCACGCTCGACCAGC 45

# RESULT 6

BE524251 223 bp mRNA EST 19-MAR-2001  
 LOCUS M47G12STM Arabidopsis developing seed Arabidopsis thaliana CDNA  
 DEFINITION clone M47G12 5', mRNA sequence.  
 ACCESSION BE524251  
 VERSION BE524251.1 GI:9782229  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 223)  
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
 A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
 Plant Physiol. 124 (4), 1582-1594 (2000)  
 20567808  
 Contact: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 , USA  
 Tel: 517 355 1609  
 Fax: 517 353 9334  
 Email: benning@msu.edu  
 Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

# REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

# FEATURES

source  
 1. .223  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="M47G12"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"

BASE COUNT 37 a 69 c 57 g 60 t  
 ORIGIN /note="Organ: Developing seed; Vector: pBluescript SK-;  
 Site\_1: EcoRI; Site\_2: XhoII"

Query Match 85.5%; Score 18.8; DB 10; Length 223;  
 Best Local Similarity 90.9%; Pred. No. 84;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 99 AACGAGTGCACGCTCGACCAGC 120

RESULT 7  
 BE523875 225 bp mRNA EST 19-MAR-2001  
 LOCUS M43A1STM Arabidopsis developing seed Arabidopsis thaliana CDNA  
 DEFINITION Clone M43A1 5', mRNA sequence.  
 ACCESSION BE523875  
 VERSION BE523875.1 GI:9781853  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 TITLE White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de  
 Illarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
 JOURNAL A new set of Arabidopsis expressed sequence tags from developing  
 MEDLINE seeds. The metabolic pathway from carbohydrates to seed oil  
 COMMENT Plant Physiol. 124 (4), 1582-1594 (2000)  
 20567808

CONTACT: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 , USA  
 Tel: 517 355 1609  
 Fax: 517 353 9334  
 Email: benning@msu.edu  
 Michigan State University DNA Sequencing Facility Arabidopsis  
 Biological Resource Center, The Ohio State University, 309 Botany &  
 Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:  
 6142920603 TEL: 6142929371.

FEATURES  
 source  
 Location/Qualifiers  
 1..225  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="M43A1"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"  
 /note="Organ: Developing seed; Vector: pBluescript SK-;  
 Site\_1: EcoRI; Site\_2: XhoII"  
 BASE COUNT 48 a 68 c 52 g 56 t  
 ORIGIN 1 others

Query Match 85.5%; Score 18.8; DB 10; Length 225;  
 Best Local Similarity 90.9%; Pred. No. 84;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 118 AACGAGTGCACGCTCGACCAGC 139

RESULT 8

BE522306 231 bp mRNA EST 19-MAR-2001  
 LOCUS M25A1STM Arabidopsis developing seed Arabidopsis thaliana CDNA  
 DEFINITION Clone M25A1 5', mRNA sequence.  
 ACCESSION BE522306  
 VERSION BE522306.1 GI:9780284  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 TITLE White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de  
 Illarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
 JOURNAL A new set of Arabidopsis expressed sequence tags from developing  
 MEDLINE seeds. The metabolic pathway from carbohydrates to seed oil  
 COMMENT Plant Physiol. 124 (4), 1582-1594 (2000)  
 20567808

CONTACT: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 , USA  
 Tel: 517 355 1609  
 Fax: 517 353 9334  
 Email: benning@msu.edu  
 Michigan State University DNA Sequencing Facility Arabidopsis  
 Biological Resource Center, The Ohio State University, 309 Botany &  
 Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:  
 6142920603 TEL: 6142929371.

FEATURES  
 source  
 Location/Qualifiers  
 1..231  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="M25A1"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"  
 /note="Organ: Developing seed; Vector: pBluescript SK-;  
 Site\_1: EcoRI; Site\_2: XhoII"  
 BASE COUNT 45 a 72 c 57 g 57 t  
 ORIGIN 1

Query Match 85.5%; Score 18.8; DB 10; Length 231;  
 Best Local Similarity 90.9%; Pred. No. 85;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22  
 ||||| ||||| ||||| ||||| |||||  
 Db 57 AACGAGTGCACGCTCGACCAGC 78

RESULT 9  
 BE520764 232 bp mRNA EST 19-MAR-2001  
 LOCUS M14H9STM Arabidopsis developing seed Arabidopsis thaliana CDNA  
 DEFINITION Clone M14H9 5', mRNA sequence.  
 ACCESSION BE520764  
 VERSION BE520764.1 GI:9778742  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 TITLE White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de  
 Illarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
 JOURNAL A new set of Arabidopsis expressed sequence tags from developing

seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Michigan State University DNA Sequencing Facility Arabidopsis  
Biological Resource Center, The Ohio State University, 309 Botany &  
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:  
6142920603 TEL: 6142929371.  
Location/Qualifiers

FEATURES  
source  
1. .232  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="M14H9"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
44 a 73 c 57 g 58 t

BASE COUNT 44 a 73 c 57 g 58 t  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 232;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22  
||||||| ||||| |||||  
Db 57 AACGAGTGCAGCTCGACCAGC 78

RESULT 10  
BE525569  
LOCUS M62M13STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
DEFINITION Clone 600014455R1 5', mRNA sequence.  
ACCESSION BE525569  
VERSION BE525569  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 232)  
AUTHORS Illarduya, O., Javorski, J.G., Ohlrogge, J. and Benning, C.  
TITLE A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
COMMENT Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.  
Location/Qualifiers

FEATURES  
source  
1. .238  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014490R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
45 a 70 c 63 g 58 t 2 others

BASE COUNT 45 a 70 c 63 g 58 t 2 others  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 238;

source  
1. .232  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014455R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
41 a 72 c 56 g 63 t

BASE COUNT 41 a 72 c 56 g 63 t  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 232;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22  
||||||| ||||| |||||  
Db 90 AACGAGTGCAGCTCGACCAGC 111

RESULT 11  
BE525596  
LOCUS M62C23STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
DEFINITION Clone 600014490R1 5', mRNA sequence.  
ACCESSION BE525596  
VERSION BE525596  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 238)  
AUTHORS White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de  
Illarduya, O., Javorski, J.G., Ohlrogge, J. and Benning, C.  
TITLE A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
COMMENT Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.  
Location/Qualifiers

FEATURES  
source  
1. .238  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014490R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
45 a 70 c 63 g 58 t 2 others

BASE COUNT 45 a 70 c 63 g 58 t 2 others  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 238;

```

Best Local Similarity 90.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
||||| ||||| ||||| |||||
Db 24 AACGAGTCCAGCTCGACCAGC 45

RESULT 12
BE523203 239 bp mRNA EST 19-MAR-2001
DEFINITION M35C7STM Arabidopsis developing seed Arabidopsis thaliana CDNA
ACCESSION BE523203
VERSION BE523203.1 GI:9781277
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 239)
REFERENCE White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
AUTHORS Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
JOURNAL seeds. The metabolic pathway from carbohydrates to seed oil
MEDLINE Plant Physiol. 124 (4), 1582-1594 (2000)
COMMENT Contact: Benning, C
20567808 Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA, FAX: 6142920603 TEL: 6142929371.
FEATURES
source
Location/Qualifiers
1..239
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="M35C7"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
BASE COUNT 49 a 74 c 57 g 59 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 239;
Best Local Similarity 90.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
||||| ||||| ||||| |||||
Db 113 AACGAGTCCAGCTCGACCAGC 134

RESULT 13
BE525564 243 bp mRNA EST 19-MAR-2001
LOCUS M62C13STM Arabidopsis developing seed Arabidopsis thaliana CDNA
DEFINITION clone 600014450R1 5', mRNA sequence.
ACCESSION BE525564
VERSION BE525564.1 GI:9783465
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 244)
REFERENCE White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
AUTHORS Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
JOURNAL seeds. The metabolic pathway from carbohydrates to seed oil
MEDLINE Plant Physiol. 124 (4), 1582-1594 (2000)
COMMENT Contact: Benning, C
20567808 Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
FEATURES
source
Location/Qualifiers
1..239
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="M35C7"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
BASE COUNT 49 a 74 c 57 g 59 t
ORIGIN

```

```

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 243)
REFERENCE White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
AUTHORS Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
JOURNAL seeds. The metabolic pathway from carbohydrates to seed oil
MEDLINE Plant Physiol. 124 (4), 1582-1594 (2000)
COMMENT Contact: Benning, C
20567808 Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA, FAX: 6142920603 TEL: 6142929371.
FEATURES
source
Location/Qualifiers
1..243
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
BASE COUNT 49 a 69 c 65 g 59 t
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 243;
Best Local Similarity 90.9%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
||||| ||||| ||||| |||||
Db 21 AACGAGTCCAGCTCGACCAGC 42

RESULT 14
BE525454 244 bp mRNA EST 19-MAR-2001
LOCUS M62G03STM Arabidopsis developing seed Arabidopsis thaliana CDNA
DEFINITION clone 600014412R1 5', mRNA sequence.
ACCESSION BE525454
VERSION BE525454.1 GI:9783432
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 244)
REFERENCE White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
AUTHORS Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
JOURNAL seeds. The metabolic pathway from carbohydrates to seed oil
MEDLINE Plant Physiol. 124 (4), 1582-1594 (2000)
COMMENT Contact: Benning, C
20567808 Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA, FAX: 6142920603 TEL: 6142929371.
FEATURES
source
Location/Qualifiers
1..243
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
BASE COUNT 49 a 69 c 65 g 59 t
ORIGIN

```

USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.

FEATURES  
source

1. .244  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="600014414R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
BASE COUNT 46 a 74 c 66 g 57 t 1 others  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 244;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22  
||||||| ||||| ||||| |||||  
Db 44 AACGAGTGCACGCTCGACCAGC 65

## RESULT 15

BE525533  
LOCUS BE525533 244 bp mRNA EST 19-MAR-2001  
DEFINITION M62K03STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
clone 600014414R1 5', mRNA sequence.  
ACCESSION BE525533  
VERSION BE525533.1 GI:9783434  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 244)  
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de  
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
USA

Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.

FEATURES  
source

1. .244  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="600014414R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"

/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
BASE COUNT 43 a 74 c 60 g 67 t  
ORIGIN  
Query Match 85.5%; Score 18.8; DB 10; Length 244;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 aacgagtgtcagctagaccagc 22  
||||||| ||||| ||||| |||||  
Db 91 AACGAGTGCACGCTCGACCAGC 112  
Search completed: February 15, 2002, 18:04:22  
Job time: 20971 sec

---



BASE COUNT 5 a 7 c 5 g 5 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 1 CGCAGTTCTGTGAACATCGACC 22

## RESULT 2

AX172482  
LOCUS AX172482 22 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 43 from Patent WO0141558.  
ACCESSION AX172482  
VERSION AX172482.1 GI:14597594  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS de Both, G. and de Beuckeleer, M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 43 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
1. .22  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer BNA06"

BASE COUNT 5 a 7 c 5 g 5 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 1 CGCAGTTCTGTGAACATCGACC 22

## RESULT 3

I24540/c  
LOCUS I24540 3113 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 20 from patent US 5543576.  
ACCESSION I24540  
VERSION I24540.1 GI:1604410  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3113)  
AUTHORS van Ooijen, A.J.J., Rietveld, K., Hoekema, A., Pen, J., Sijmons, P.C., Verwoerd, T.C. and Quax, W.J.  
TITLE production of enzymes in seeds and their use  
JOURNAL Patent: US 5543576-A 20 06-AUG-1996;  
FEATURES  
source  
1. .3113  
/organism="unknown"  
BASE COUNT 961 a 685 c 586 g 881 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22

Db 1183 CGCAGTTCTGTGAACATCGACC 1162  
|||||

## RESULT 4

I83673/c  
LOCUS I83673 3113 bp DNA PAT 10-AUG-1998  
DEFINITION Sequence 2 from patent US 5714474.  
ACCESSION I83673  
VERSION I83673.1 GI:3407203  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3113)  
AUTHORS van Ooijen, A.J.J., Rietveld, K., Hoekema, A., Pen, J., Sijmons, P., Christman, Verwoerd, T., Cornelis and Quax, W. Johannes.  
TITLE Production of enzymes in seeds and their use  
JOURNAL Patent: US 5714474-A 2 03-FEB-1998;  
FEATURES  
source  
1. .3113  
/organism="unknown"  
BASE COUNT 961 a 684 c 586 g 882 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 1183 CGCAGTTCTGTGAACATCGACC 1162

## RESULT 5

BNCRU4/c  
LOCUS BNCRU4 3113 bp DNA PLN 10-FEB-1999  
DEFINITION Brassica napus cruA gene for cruciferin.  
ACCESSION X14555  
VERSION X14555.1 GI:17810  
KEYWORDS cruA gene; cruciferin; seed storage protein.  
SOURCE rape.  
ORGANISM Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 3113)  
AUTHORS Ryan, A.J.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAR-1989) Ryan A.J., Department of Biological Sciences, University of Durham, South Road, Durham, DH1 3LE, United Kingdom  
REFERENCE 2 (bases 1 to 3113)  
AUTHORS Ryan, A.J., Royal, C.L., Hutchinson, J. and Shaw, C.H.  
TITLE Genomic sequence of a 12S seed storage protein from oilseed rape (Brassica napus c.v. Jet Neuf)  
JOURNAL Nucleic Acids Res. 17 (9), 3584 (1989)  
MEDLINE 89263796  
FEATURES  
source  
1. .3113  
Location/Qualifiers  
/organism="Brassica napus"  
/strain="Jet Neuf"  
/db\_xref="taxon:3708"  
/tissue\_type="leaf"  
644..650  
/note="pot. TATA-box"  
680..3010  
precursor\_RNA  
CDS  
/note="primary transcript"  
join(709..991,1220..1581,2049..2870)  
/codon\_start=1  
/product="cruciferin"  
/protein\_id="CAA32692.1"



```
/db_xref="GI:17811"
/db_xref="SWISS-PROT:P11090"
/translation="MARLSLLSFLALLFLHGTAAQFPNECOLQNALPEPSHVL
KAEAGRIEVDHHAHAPOLRCGVSFVRYIIESKGLYLPSFFSTAKLSFVAKGEGLMGRV
VLCATFEODSSVFOPSGGSPFEGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOG
FRDMHQKVEHIRTGTDTIATHPGVAQWFYNDGNQPLVIVSYLDLASHQNLDRNPRFF
LAGNPOQGVWIEGREQOQKNILNGFTPEVLAKAFKIDVTAQQLQNOQDNRGNII
VQGPFSVIRPPLRSORPOETEYNGLEETICSACTDNLDLDDPSNADVYKPOLGYISTL
YDPLILRLSALRGSIRQANAVLPQWNNANAVLYVTDGEAHQVQVNDNGDRVFDG
QVSOQLLSIPOGFSVVKRATSEQFRWIEFTKTNANAQINTLAGRTSVLRGLPLEVISN
GYOISLEEARVRKENTIEITLTHSSGPASYGGRPKADA"
intron 992..1219
/note="intron I"
intron 1582..2048
/note="intron II"
misc_feature 2915..2920
/note="polyA signal"
misc_feature 2989..2994
/note="polyA signal"
BASE COUNT 961 a 685 c 586 g 881 t
ORIGIN
Query Match 100.0%; Score 22; DB 8; Length 3113;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgcagttctgtgaacatcgacc 22
|||||
Db 1183 CGCAGTTCGTGAACATCGACC 1162

RESULT 6
BNC1G/c BNC1G 3198 bp DNA PLN 04-APR-1995
LOCUS B napus Bnc1 gene for cruciferin storage protein.
ACCESSION X59294
VERSION X59294.1 GI:17790
KEYWORDS cruciferin; cruciferin storage protein.
SOURCE rape.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 3198)
Breen,J.P. and Crouch,M.L.
Molecular analysis of a cruciferin storage protein gene family of
Brassica napus
Plant Mol. Biol. 19 (6), 1049-1055 (1992)
92379259
2 (bases 1 to 3198)
Breen,J.P.
Direct Submission
Submitted (24-APR-1991) Breen J.P., Jordan Hall, Indiana
University, Bloomington, IN 47401, USA
See also M16860 & X59295 (for Bnc2 gene).
FEATURES
Location/Qualifiers
1..3198
/organism="Brassica napus"
/strain="C.V. Tower"
/db_xref="taxon:3708"
/cell_line="ED8767"
/clone_lib="Ch4a"
/clone="lambda Bnc1"
Join(<709..991,1220..1584,2069..2488,2653..>3057)
/gene="Bnc1"
709..3057
/gene="Bnc1"
Join(709..991,1220..1584,2069..2488,2653..3057)
/gene="Bnc1"
/codon_start=1
/product="cruciferin storage protein"
/protein_id="CAA41984.1"
```

```
/db_xref="GI:762919"
/db_xref="SWISS-PROT:P33523"
/translation="MARLSLLSFLALLFLHGTAAQFPNECOLQNALPEPSHVL
KAEAGRIEVDHHAHAPOLRCGVSFVRYIIESKGLYLPSFFSTAKLSFVAKGEGLMGRV
VLCATFEODSSVFOPSGGSPFEGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOG
FRDMHQKVEHIRTGTDTIATHPGVAQWFYNDGNQPLVIVSYLDLASHQNLDRNPRFF
LAGNPOQGVWIEGREQOQKNILNGFTPEVLAKAFKIDVTAQQLQNOQDNRGNII
VQGPFSVIRPPLRSORPOETEYNGLEETICSACTDNLDLDDPSNADVYKPOLGYISTL
YDPLILRLSALRGSIRQANAVLPQWNNANAVLYVTDGEAHQVQVNDNGDRVFDG
QVSOQLLSIPOGFSVVKRATSEQFRWIEFTKTNANAQINTLAGRTSVLRGLPLEVI
SNQYQISLEEARVRKENTIEITLTHSSGPASYGGRPKADA"
<709..991
/gene="Bnc1"
/number=1
992..1219
/gene="Bnc1"
/number=1
1220..1584
/gene="Bnc1"
/number=2
1585..2068
/gene="Bnc1"
/number=2
2069..2488
/gene="Bnc1"
/number=3
2489..2652
/gene="Bnc1"
/number=3
2654..>3057
/gene="Bnc1"
/number=4
polyA_signal 3102..3107
polyA_signal 3176..3181
BASE COUNT 992 a 690 c 600 g 916 t
ORIGIN
Query Match 100.0%; Score 22; DB 8; Length 3198;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgcagttctgtgaacatcgacc 22
|||||
Db 1183 CGCAGTTCGTGAACATCGACC 1162

RESULT 7
AC021599
ID AC021599 standard; DNA; HTG; 143830 BP.
XX AC021599;
XX AC021599;
XX AC021599.3
XX 17-JAN-2000 (Rel. 62, Created)
DT 21-SEP-2000 (Rel. 65, Last updated, Version 3)
XX Homo sapiens clone RP11-189E18, WORKING DRAFT SEQUENCE, 29 unordered
DE pieces.
DE HTG; HTGS_DRAFT; HTGS_PHASE1.
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX [1]
RN 1-143830
RP Birren B., Linton L., Nusbaum C., Lander E.;
RA "Homo sapiens, clone RP11-189E18";
RT Unpublished.
RL XX
```



FT misc\_feature /note="assembly\_fragment"  
 FT 31237..35355  
 FT /note="assembly\_fragment"  
 FT 35456..39501  
 FT /note="assembly\_fragment"  
 FT 39602..44277  
 FT /note="assembly\_fragment"  
 FT 44378..47808  
 FT /note="assembly\_fragment"  
 FT 47909..52253  
 FT /note="assembly\_fragment"  
 FT 52354..57551  
 FT /note="assembly\_fragment"  
 FT 57652..64108  
 FT /note="assembly\_fragment"  
 FT 64209..71334  
 FT /note="assembly\_fragment"  
 FT 71435..79411  
 FT /note="assembly\_fragment"  
 FT 79512..86611  
 FT /note="assembly\_fragment"  
 FT 86712..94930  
 FT /note="assembly\_fragment"  
 FT 95031..101723  
 FT /note="assembly\_fragment"  
 FT 101824..116518  
 FT /note="assembly\_fragment"  
 FT 116619..129081  
 FT /note="assembly\_fragment"  
 FT 129182..143830

Query Match 79.1%; Score 17.4; DB 33; Length 143830;  
 Best Local Similarity 94.7%; Pred. No. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20  
 |||||  
 Db 83635 GCAGTTCTGTGAACATAGA 83653

RESULT 8  
 AC092613/c  
 LOCUS AC092613 151210 bp DNA HTG 19-JUL-2001  
 DEFINITION Homo sapiens chromosome 7 clone RP11-189E18, WORKING DRAFT  
 SEQUENCE, 1 unordered pieces.  
 ACCESSION AC092613 AC021599  
 VERSION AC092613.1 GI:14916198  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 151210)  
 Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 151210)  
 Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-JUL-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 COMMENT On Jul 19, 2001 this sequence version replaced gi:10198499.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc/index.shtml>  
 ----- Project Information -----  
 Center project name: H\_NH0189E18  
 Drafting center: WIBR  
 ----- Summary Statistics -----

Sequencing vector: M13; 26%  
 Chemistry: Dye-primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 63% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 148642 bases at least Q40  
 Consensus quality: 149695 bases at least Q30  
 Consensus quality: 151204 bases at least Q20  
 Insert size: 143000; agarose-fp  
 Quality coverage: 8.04 in Q20 bases; agarose-fp  
 Quality coverage: 7.60 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 151210: contig of 151210 bp in length.

FEATURES  
 source  
 1..151210  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /clone="RP11-189E18"  
 1..151210  
 /note="assembly\_name:Contig5  
 clone\_end:SP6  
 vector\_side:right"  
 BASE COUNT 45173 a 28928 c 30068 g 47041 t  
 ORIGIN

Query Match 79.1%; Score 17.4; DB 2; Length 151210;  
 Best Local Similarity 94.7%; Pred. No. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20  
 |||||  
 Db 17340 GCAGTTCTGTGAACATAGA 17322

RESULT 9  
 AC005047  
 LOCUS AC005047 154016 bp DNA PRI 07-OCT-2000  
 DEFINITION Homo sapiens BAC clone CTB-14E15 from 7q22-q32, complete sequence.  
 ACCESSION AC005047  
 VERSION AC005047.3 GI:10716665  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 154016)  
 Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 REFERENCE 2 (bases 1 to 154016)  
 Courtney, L., Gillam, B., Stoneking, T., Elliott, G. and Langston, Y.  
 TITLE The sequence of Homo sapiens BAC clone CTB-14E15  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 154016)  
 Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 4 (bases 1 to 154016)  
 Waterston, R.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (07-OCT-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 7, 2000 this sequence version replaced gi:7631121.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
Center project name: H\_RG014E15  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30);  
an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send  
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-14E15 is from the first release of the human BAC library  
CTB-978SK-B. The library contains cloned DNA from the male  
fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl.  
Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8  
(1996). This clone is available from Research Genetics, Inc.  
(<http://www.resgen.com>).  
VECTOR: pHELOBAC11  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is GS1-34D21, 200 base pair  
overlap. Actual start of this clone is at base position 1 of  
CTB-14E15; actual end is at base position 7330 of GS1-34D21.

Due to low quality, the fidelity of the sequence from 130932 to  
130982 can not be guaranteed.

FEATURES  
source

Location/Qualifiers  
1..154016  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q22-q32"  
/clone="CTB-14E15"  
/clone\_lib="CTB-978SK-B"  
589..874  
/rpt\_family="L1"  
903..2491  
/rpt\_family="ERV1"  
2495..2700  
/rpt\_family="L1"  
2720..2838  
/rpt\_family="Alu"  
2839..3289  
/rpt\_family="L1"  
3411..3476  
/rpt\_family="MaLR"

repeat\_region 3477..3783  
/rpt\_family="Alu"  
repeat\_region 3784..4135  
/rpt\_family="MaLR"  
repeat\_region 4849..5153  
/rpt\_family="L1"  
repeat\_region 5740..6039  
/rpt\_family="Alu"  
repeat\_region 6077..6151  
/rpt\_family="MER99"  
repeat\_region 6319..6616  
/rpt\_family="CR1"  
repeat\_region 8583..8761  
/rpt\_family="Alu"  
repeat\_region 8771..8906  
/rpt\_family="Alu"  
repeat\_region 9238..9514  
/rpt\_family="Alu"  
repeat\_region 10478..10731  
/rpt\_family="ERV1"  
STS 11164..11523  
/db\_xref="GI:6120582"  
repeat\_region 12056..12181  
/rpt\_family="MaLR"  
repeat\_region 12256..12482  
/rpt\_family="MIR"  
repeat\_region 12500..12625  
/rpt\_family="CR1"  
repeat\_region 12717..12750  
/rpt\_family="CR1"  
repeat\_region 12817..12901  
/rpt\_family="CR1"  
repeat\_region 13158..13555  
/rpt\_family="ERV1"  
repeat\_region 14352..14508  
/rpt\_family="L1"  
repeat\_region 14509..14900  
/rpt\_family="MaLR"  
repeat\_region 14901..15114  
/rpt\_family="L1"  
repeat\_region 15152..15625  
/rpt\_family="ERV1"  
repeat\_region 15626..16096  
/rpt\_family="L1"  
repeat\_region 16147..16636  
/rpt\_family="ERV1"  
repeat\_region 16640..17690  
/rpt\_family="ERVK"  
repeat\_region 17710..17806  
/rpt\_family="ERV1"  
repeat\_region 17850..17908  
/rpt\_family="L2"  
repeat\_region 18099..18221  
/rpt\_family="MER103"  
STS 18305..18622  
/db\_xref="GI:485206"  
repeat\_region 18833..19179  
/rpt\_family="MER2\_type"  
repeat\_region 19215..19509  
/rpt\_family="Alu"  
repeat\_region 19983..20033  
/rpt\_family="L2"  
repeat\_region 20109..20370  
/rpt\_family="L2"  
repeat\_region 20708..20760  
/rpt\_family="MaLR"  
repeat\_region 20933..21089  
/rpt\_family="MaLR"  
repeat\_region 21094..21147  
/rpt\_family="ERV1"  
repeat\_region 21148..21458  
/rpt\_family="Alu"  
repeat\_region 21459..21613

```

repeat_region      /rpt_family="ERV1"
21626..21980
/rpt_family="MaLR"
21985..22223
/rpt_family="ERV1"
22241..22547
/rpt_family="Alu"
22613..22860
/rpt_family="MaLR"
22919..23389
/rpt_family="ERV1"
23390..23679
/rpt_family="Alu"
23680..23805
/rpt_family="ERV1"
24361..24460
/rpt_family="L2"
24491..24642
/rpt_family="MIR"
24643..25009
/rpt_family="MaLR"
25010..25099
/rpt_family="MIR"
25230..25647
/rpt_family="MaLR"
25939..26236
/rpt_family="Alu"
26723..29014

```

Query Match 79.1%; Score 17.4; DB 9; Length 154016;  
 Best Local Similarity 94.7%; Pred. No. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20  
 ||||| ||||| ||||| ||

Db 67229 GCAGTTCTGTGACATAGA 67247

```

RESULT 10
AF142615/c AF142615 287 bp DNA MAM 22-JUN-2000
LOCUS Equus caballus microsatellite COR078 sequence.
DEFINITION AF142615
ACCESSION AF142615.1 GI:6318588
VERSION AF142615.1
KEYWORDS horse.
SOURCE Equus caballus
ORGANISM Equus caballus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
1 (bases 1 to 287)
Tallmadge,R.L., Hopman,T.J., Schug,M.D., Aquadro,C.F.,
Bowling,A.T., Murray,J.D., Caetano,A.R. and Antczak,D.F.
TITLE Equine dinucleotide repeat loci COR061-COR080
JOURNAL Anim. Genet. 30 (6), 462-463 (1999)
MEDLINE 20078362
PUBMED 10612238
REFERENCE 2 (bases 1 to 287)
AUTHORS Tallmadge,R.L., Hopman,T.J., Schug,M.D., Aquadro,C.F.,
Bowling,A.T., Murray,J.D., Caetano,A.R. and Antczak,D.F.
TITLE Direct Submission
JOURNAL Submitted (13-APR-1999) Cornell University, J.A. Baker Institute,
Hungerford Hill Rd., Ithaca, NY 14853, USA
FEATURES
source
1..287
/organism="Equus caballus"
/db_xref="taxon:9796"
/chromosome="2"
59..70
/note="microsatellite COR078"
/rpt_type=tandem
/rpt_unit=gt
BASE COUNT 56 a 37 c 102 g 92 t
ORIGIN

```

```

Query Match 78.2%; Score 17.2; DB 4; Length 287;
Best Local Similarity 86.4%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
||||| ||||| ||||| ||
Db 201 CGCAGTTCTGTGCACATGGTCC 180

RESULT 11
I56095 I56095 587 bp DNA PAT 07-OCT-1997
LOCUS Sequence 3 from patent US 5648481.
DEFINITION I56095
ACCESSION I56095
VERSION I56095.1 GI:2476889
KEYWORDS .
SOURCE Unknown.
ORGANISM Unclansified.
REFERENCE 1 (bases 1 to 587)
AUTHORS Parodos,K. and McCarty,J.
TITLE Nucleic acid probes for the detection of shigella
JOURNAL Patent: US 5648481-A 3 15-JUL-1997;
FEATURES Location/Qualifiers
1..587
source /organism="unknown"
BASE COUNT 138 a 149 c 139 g 161 t
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 587;
Best Local Similarity 86.4%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
||||| ||||| ||||| ||
Db 566 CGCAGTACTGTGAACCTCGATC 587

RESULT 12
AF190914 AF190914 646 bp DNA BCT 16-JAN-2000
LOCUS Escherichia coli strain K2 eltB gene, 3' flanking sequence.
DEFINITION AF190914
ACCESSION AF190914
VERSION AF190914.1 GI:6694242
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Plasmid Escherichia coli
REFERENCE Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
AUTHORS Escherichia.
1 (bases 1 to 646)
Schloer,S., Riedl,S., Blass,J. and Riedl,J.
TITLE Genetic rearrangements of the regions adjacent to genes encoding
heat-labile enterotoxins (eltAB) of enterotoxigenic Escherichia
coli strains
JOURNAL Appl. Environ. Microbiol. 66 (1), 352-358 (2000)
MEDLINE 20087561
PUBMED 10618247
REFERENCE 2 (bases 1 to 646)
AUTHORS Riedl,J., Schloer,S., Riedl,S. and Blass,J.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-1999) Zentrum fuer Infektionsforschung,
University of Wuerzburg, Roentgenring 11, Wuerzburg 97070, Germany
FEATURES Location/Qualifiers
1..646
/organism="Escherichia coli"
/plasmid="Ent"
/strain="K2"
/db_xref="taxon:562"
<1..>646
/gene="eltB"
source
gene

```

BASE COUNT 182 a 134 c 149 g 175 t 6 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 1; Length 646;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 586 CGCAGTACTGTGAACCTCGATC 607

RESULT 13

LOCUS I56097 1174 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 5 from patent US 5648481.  
ACCESSION I56097  
VERSION I56097.1 GI:2476891

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

AUTHORS 1 (bases 1 to 1174)

TITLE Parodos,K. and McCarty,J.

JOURNAL Nucleic acid probes for the detection of shigella

FEATURES Patent: US 5648481-A 5 15-JUL-1997;

Location/Qualifiers

1..1174

source /organism="unknown"

BASE COUNT 252 a 296 c 276 g 330 t 20 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 1174;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 322 CGCAGTACTGTGAACCTCGATC 343

RESULT 14

LOCUS I56101 1188 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 9 from patent US 5648481.  
ACCESSION I56101

VERSION I56101.1 GI:2476895

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

AUTHORS 1 (bases 1 to 1188)

TITLE Parodos,K. and McCarty,J.

JOURNAL Nucleic acid probes for the detection of shigella

FEATURES Patent: US 5648481-A 9 15-JUL-1997;

Location/Qualifiers

1..1188

source /organism="unknown"

BASE COUNT 258 a 311 c 273 g 330 t 16 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 1188;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 15

LOCUS I56099 1196 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 7 from patent US 5648481.  
ACCESSION I56099

VERSION I56099.1 GI:2476893

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

AUTHORS 1 (bases 1 to 1196)

TITLE Parodos,K. and McCarty,J.

JOURNAL Nucleic acid probes for the detection of shigella

FEATURES Patent: US 5648481-A 7 15-JUL-1997;

Location/Qualifiers

1..1196

source /organism="unknown"

BASE COUNT 259 a 298 c 272 g 353 t 14 others  
ORIGIN

Query Match

Best Local Similarity 78.2%; Score 17.2; DB 6; Length 1196;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22

||||| ||||||| |||||

Db 195 CGCAGTACTGTGAACCTCGATC 216

Search completed: February 15, 2002, 18:55:41  
Job time: 20580 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:02:07 ; Search time 868.33 Seconds  
(without alignments)  
21.721 Million cell updates/sec

Title: US-09-698-903B-14

Perfect score: 22  
Sequence: 1 cgcagttctggaacatcgacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	22	100.0	22	22	AAH25458	PCR primer for end
2	22	100.0	22	22	AAH25458	PCR primer B04, to
3	22	100.0	3113	12	AAQ07003	Cruciferin A gene.
4	17.2	78.2	587	14	AAQ37508	Shigella specific
5	17.2	78.2	1174	14	AAQ37510	E. coli repeat 1.
6	17.2	78.2	1188	14	AAQ37514	S. flexneri repeat
7	17.2	78.2	1196	14	AAQ37512	E. coli repeat 2
8	16.8	76.4	2019	19	AAV38672	Mus musculus SOCS7
9	16.8	76.4	7316	22	AAH21115	C. glutamicum LP-6
10	16.4	74.5	4961	21	AAC77036	Human ORFX ORF2591
11	15.8	71.8	226	21	AAC12754	Human secreted pro

c	12	15.8	71.8	319608	21	AAH51601	Human chromosome 1
c	13	15.8	71.8	319608	22	AAH51601	Human chromosome 1
	14	15.6	70.9	545	20	AAH24719	Tobacco leaf polyp
	15	15.6	70.9	769	18	AAH30779	Streptococcus pneu
	16	15.6	70.9	769	19	AAH28688	DNA encoding a S.
	17	15.6	70.9	2108	22	AAH56464	Neisseria meningit
	18	15.6	70.9	2109	22	AAH56463	Neisseria meningit
	19	15.6	70.9	2111	22	AAH56458	Neisseria meningit
	20	15.6	70.9	2112	21	AAH5745	DNA encoding a Nei
	21	15.6	70.9	2112	21	AAH54329	Neisseria meningit
	22	15.6	70.9	2112	22	AAH56459	Neisseria meningit
	23	15.6	70.9	2112	22	AAH56460	Neisseria meningit
	24	15.6	70.9	2112	22	AAH56461	Neisseria meningit
	25	15.6	70.9	2112	22	AAH56462	Neisseria meningit
	26	15.6	70.9	2112	22	AAH56466	Neisseria meningit
	27	15.6	70.9	2112	22	AAH56467	Neisseria meningit
c	28	15.6	70.9	2582	7	AAH60610	Vector sequence of
c	29	15.6	70.9	2585	8	AAH70629	DNA a sequence of
c	30	15.6	70.9	8367	21	AAH81483	N. meningitidis pa
	31	15.6	70.9	21185	21	AAH63350	Streptomyces globi
	32	15.6	70.9	63164	21	AAH63348	Streptomyces globi
	33	15.6	70.9	349980	21	AAH21544	Neisseria meningit
	34	15.6	70.9	349980	21	AAH21607	Neisseria meningit
	35	15.6	70.9	1437668	21	AAH81490	N. meningitidis B
c	36	15.4	70.0	140	21	AAH06575	Human secreted pro
c	37	15.4	70.0	256	21	AAH15641	Human prostate can
c	38	15.4	70.0	429	22	AAH14743	Probe #4676 for ge
c	39	15.4	70.0	429	22	AAH36109	Probe #4795 used t
c	40	15.4	70.0	429	22	AAH04546	Probe #4537 used t
	41	15.4	70.0	947	22	AAH55404	Nucleotide sequenc
	42	15.4	70.0	947	22	AAH55595	Nucleotide sequenc
	43	15.4	70.0	947	22	AAH57184	Apoptin-associatin
	44	15.4	70.0	1131	22	AAH55405	Nucleotide sequenc
	45	15.4	70.0	1131	22	AAH55596	Nucleotide sequenc

ALIGNMENTS

RESULT 1  
AAH25458  
ID AAH25458 standard; DNA; 22 BP.  
XX AC  
XX AAH25458;  
XX AC  
DT 05-SEP-2001 (first entry)  
XX PCR primer for endogenous sequences in transgenic plants.

DE Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
KW fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.  
XX Synthetic.  
OS

XX WO200141558-Al.  
XX 14-JUN-2001.  
XX

XX 06-DEC-2000; 2000WO-EPI2872.  
XX 08-DEC-1999; 99US-0457037.  
XX

XX (AVET ) AVENTIS CROPS SCIENCE NV.  
XX De Both G, De Beuckeleer M;  
XX WPI; 2001-381419/40.  
XX

PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
PT with improved qualities, comprises a male-sterility gene and fertility  
PT restorer gene, integrated into the genome -  
XX Example 5; Page 53; 98pp; English.  
PS

XX

CC The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAH25457-58 were used to amplify endogenous sequences  
 CC from transgenic plants of the invention.

XX

SQ Sequence 22 BP; 5 A; 7 C; 5 G; 5 T; 0 other;

## Query Match

Best Local Similarity 100.0%; Score 22; DB 22; Length 22;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 1 cgcagttctgtgaacatcgacc 22  
 |||||

## RESULT 2

AA07003  
 ID AAD07003 standard; DNA; 22 BP.

XX AAD07003;

XX 06-AUG-2001 (first entry)

DE PCR primer B04, to recognise foreign DNA and flanking sequence of MS-B2.

KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; ss.

XX

XX Brassica napus.

XX WO200131042-A2.

XX 03-MAY-2001.

PF 26-OCT-2000; 2000WO-EP10680.  
 PR 29-OCT-1999; 99US-0430497.

XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA Weston B, De Beuckeleer M;  
 PI WPI; 2001-300517/31.

XX

DR Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome.  
 XX Example 5; Page 33; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.

CC The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.  
 XX Sequence 22 BP; 5 A; 7 C; 5 G; 5 T; 0 other;

XX

## Query Match

Best Local Similarity 100.0%; Score 22; DB 22; Length 22;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 1 cgcagttctgtgaacatcgacc 22  
 |||||

## RESULT 3

AAQ13870/C  
 ID AAQ13870 standard; DNA; 3113 BP.

XX AAQ13870;

DT 09-DEC-1991 (first entry)  
 DE Cruciferin A gene.

XX Seed storage protein; cruA; ss.

XX Brassica napus.

XX EP449376-A.

XX 02-OCT-1991.

XX 25-MAR-1991; 91EP-0200688.

XX 25-MAR-1991; 91EP-0200688.

XX 23-MAR-1990; 90US-0498561.

XX (KONN ) GIST-BROCADES NV.

XX Pen J, Sijmons PC, Van Ooyen AJJ, Rietveld K, Verwoerd TC;  
 PI Quax WJ;  
 XX WPI; 1991-289815/40.
XX Seeds contg. enhanced enzyme levels from transgenic plants - used  
 PT for catalysing reactions, increasing nutritional values or  
 PT treating digestive disorders.  
 XX Example; Fig 3; 38pp; English.

XX The DNA is the genomic sequence of the seed storage protein gene  
 CC cruciferin A (cruA). It can be used in the prodn. of transgenic  
 CC plants expressing cruciferin in its seeds for use in an industrial  
 CC process. The seeds contg. the cruciferin can be used without the  
 CC need for first extracting and/or isolating the enzymes. The use of  
 CC seeds for the storage of cruciferin provides a stable vehicle which  
 CC is easily packaged and transported and easily handled during use.  
 CC See also AAQ13871-QL3877.

XX Sequence 3113 BP; 961 A; 685 C; 586 G; 881 T; 0 other;

## Query Match

Best Local Similarity 100.0%; Score 22; DB 12; Length 3113;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 1183 CGCAGTTCTGTGAACATCGACC 1162  
 |||||

## RESULT 4

AAQ37508  
 ID AAQ37508 standard; DNA; 587 BP.

XX AAQ37508;

XX 17-JUN-1993 (first entry)



```

XX DE Shigella specific fragment NT15.
XX
XX KW Chromosome; Shigella; sonnei; probe; Enteroinvasive E. coli; EIEC;
XX KW virulence plasmid; detection; dysentery; ss.
XX OS Shigella sonnei.
XX
XX PN WO9303187-A.
XX
XX PD 18-FEB-1993.
XX
XX PF 28-JUL-1992; 92WO-US06617.
XX
XX PR 31-JUL-1991; 91US-0738800.
XX
XX PA (STAD ) AMOCO CORP.
XX
XX PI McCarty JM, Parodos K;
XX
XX DR WPI; 1993-076542/09.
XX
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
XX
XX Claim 3; Page 89-90; 129pp; English.
XX
XX CC The sequences given in AAQ37506-09 fragments which were derived from
XX CC the chromosome of Shigella sonnei. These fragments were used to
XX CC design probes which are specific to Shigella and Enteroinvasive E.
XX CC coli (EIEC). The probes are specific to a stable region of the
XX CC Shigella or E. coli genome, and not the unstable virulence plasmid,
XX CC which means that they are more reliable than previous probes in the
XX CC detection of dysentery causing microbes. See also AAQ37506-35.
XX
XX SQ Sequence 587 BP; 138 A; 149 C; 139 G; 161 T; 0 other;

Query Match 78.2%; Score 17.2; DB 14; Length 587;
Best Local Similarity 86.4%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22
   ||||| ||||| |||||
Db 566 cgcagtactgtgaacctcgatc 587

RESULT 5
AAQ37510
ID AAQ37510 standard; DNA; 1174 BP.
XX
XX AC AAQ37510;
XX
XX DT 17-JUN-1993 (first entry)
XX
XX DE E. coli repeat 1.
XX
XX KW Chromosome; Shigella; sonnei; flexneri; probe; Enteroinvasive; EIEC;
XX KW E. coli; transposable element; virulence plasmid; detection; dysentery;
XX KW repeat; ss.
XX
XX OS Escherichia coli.
XX
XX PN WO9303187-A.
XX
XX PD 18-FEB-1993.
XX
XX PF 28-JUL-1992; 92WO-US06617.
XX
XX PR 31-JUL-1991; 91US-0738800.
XX
XX

```

```

PA (STAD ) AMOCO CORP.
XX
XX PI McCarty JM, Parodos K;
XX
XX DR WPI; 1993-076542/09.
XX
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
XX
XX PS Disclosure; Page 91; 129pp; English.
XX
XX CC The sequences given in AAQ37510-14 are fragments which represent repeat
XX CC sequences derived from the chromosome of E. coli and S. flexneri. The
XX CC repeat is highly conserved and has characteristics of a transposable
XX CC element. Over 20 copies of the repeat sequence are found in the
XX CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1
XX CC to 3 copies in some E. coli competitors, but not in other bacterial
XX CC species. These fragments were used to design probes which are
XX CC specific to Shigella and Enteroinvasive E. coli (EIEC). The probes
XX CC are specific to a stable region of the Shigella or E. coli genome, and
XX CC not the unstable virulence plasmid, which means that they are more
XX CC reliable than previous probes in the detection of dysentery causing
XX CC microbes. See also AAQ37506-35.
XX
XX SQ Sequence 1174 BP; 252 A; 295 C; 277 G; 330 T; 20 other;

Query Match 78.2%; Score 17.2; DB 14; Length 1174;
Best Local Similarity 86.4%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22
   ||||| ||||| |||||
Db 322 cgcagtactgtgaacctcgatc 343

RESULT 6
AAQ37514
ID AAQ37514 standard; DNA; 1188 BP.
XX
XX AC AAQ37514;
XX
XX DT 17-JUN-1993 (first entry)
XX
XX DE S. flexneri repeat 2.
XX
XX KW Chromosome; Shigella; sonnei; flexneri; probe; Enteroinvasive; EIEC;
XX KW E. coli; transposable element; virulence plasmid; detection; dysentery;
XX KW repeat; ss.
XX
XX OS Shigella flexneri.
XX
XX PN WO9303187-A.
XX
XX PD 18-FEB-1993.
XX
XX PF 28-JUL-1992; 92WO-US06617.
XX
XX PR 31-JUL-1991; 91US-0738800.
XX
XX PA (STAD ) AMOCO CORP.
XX
XX PI McCarty JM, Parodos K;
XX
XX DR WPI; 1993-076542/09.
XX
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
XX

```

PS Disclosure; Page 93; 129pp; English.

CC The sequences given in AAQ37510-14 are fragments which represent repeat  
 CC sequences derived from the chromosome of *E. coli* and *S. flexneri*. The  
 CC repeat is highly conserved and has characteristics of a transposable  
 CC element. Over 20 copies of the repeat sequence are found in the  
 CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1  
 CC to 3 copies in some *E. coli* competitors, but not in other bacterial  
 CC species. These fragments were used to design probes which are  
 CC specific to Shigella and Enteroinvasive *E. coli* (EIEC). The probes  
 CC are specific to a stable region of the Shigella or *E. coli* genome, and  
 CC not the unstable virulence plasmid, which means that they are more  
 CC reliable than previous probes in the detection of dysentery causing  
 CC microbes. See also AAQ37506-35.

XX Sequence 1188 BP; 258 A; 311 C; 273 G; 330 T; 16 other;

Query Match 78.2%; Score 17.2; DB 14; Length 1188;  
 Best Local Similarity 86.4%; Pred. No. 15;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
 ||||| ||||| ||||| |||||  
 Db 195 cgcagtactgtgaacctcgatc 216

#### RESULT 7

AAQ37512  
 ID AAQ37512 standard; DNA; 1196 BP.

AC AAQ37512;

DT 17-JUN-1993 (first entry)

DE *E. coli* repeat 2 (2).

KW Chromosome; Shigella; sonnei; flexneri; probe; Enteroinvasive; EIEC;  
 KW *E. coli*; transposable element; virulence plasmid; detection; dysentery;  
 KW repeat; ss.

OS Escherichia coli.

PN WO9303187-A.

PD 18-FEB-1993.

PF 28-JUL-1992; 92WO-US06617.

PR 31-JUL-1991; 91US-0738800.

PA (STAD ) AMOCO CORP.

PI McCarty JM, Parodos K;

DR WPI; 1993-076542/09.

XX Nucleic acid probes for detection of shigella and other pathogens  
 PT - used to diagnose dysentery in non-isotopic test format and have  
 PT utility in non-isotopic test formats requiring amplification for  
 PT high sensitivity

PS Disclosure; Page 92; 129pp; English.

CC The sequences given in AAQ37510-14 are fragments which represent repeat  
 CC sequences derived from the chromosome of *E. coli* and *S. flexneri*. The  
 CC repeat is highly conserved and has characteristics of a transposable  
 CC element. Over 20 copies of the repeat sequence are found in the  
 CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1  
 CC to 3 copies in some *E. coli* competitors, but not in other bacterial  
 CC species. These fragments were used to design probes which are  
 CC specific to Shigella and Enteroinvasive *E. coli* (EIEC). The probes  
 CC are specific to a stable region of the Shigella or *E. coli* genome, and

CC not the unstable virulence plasmid, which means that they are more  
 CC reliable than previous probes in the detection of dysentery causing  
 CC microbes. See also AAQ37506-35.

XX Sequence 1196 BP; 259 A; 298 C; 272 G; 353 T; 14 other;

Query Match 78.2%; Score 17.2; DB 14; Length 1196;  
 Best Local Similarity 86.4%; Pred. No. 15;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
 ||||| ||||| ||||| |||||  
 Db 195 cgcagtactgtgaacctcgatc 216

#### RESULT 8

AAV38672

ID AAV38672 standard; DNA; 2019 BP.

AC AAV38672;

DT 27-OCT-1998 (first entry)

DE Mus musculus SOCS7 gene.

KW SOCS; suppressor of cytokine signalling; PCR primer;  
 KW autoimmune disease; diagnosis; cancer; treatment;  
 KW cytokine mediated cellular responsiveness; hyperimmunity;  
 KW immunosuppression; allergies; hypertension; ss.

OS Mus musculus.

XX Key

FT Location/Qualifiers  
 FT 2..1054

FT /\*tag= a

FT /product= SOCS7 protein

XX WO9820023-A1.

XX 14-MAY-1998.

PF 31-OCT-1997; 97WO-AU00729.

PR 14-FEB-1997; 97AU-0005117.

PR 01-NOV-1996; 96AU-0003384.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Alexander WS, Hilton DJ, Metcalf D, Nicholson SE;  
 PI Nicola NA, Richardson RT, Starr R, Viney EM, Willson TA;  
 DR WPI; 1998-286854/25.

DR P-PSDB; AAW62621.

XX Suppressor of cytokine signalling proteins - useful to treat  
 PT disease, injury or abnormality involving cytokine mediated cellular  
 PT responsiveness e.g. hyperimmunity, immunosuppression, allergies and  
 PT hypertension

PS Claim 14; Page 143-144; 325pp; English.

XX The sequence is that of a gene encoding a suppressor of cytokine  
 CC signalling protein (SOCS). SOCS can be used to screen for naturally  
 CC occurring antibodies to SOCS, which may occur, e.g. in some autoimmune  
 CC diseases. Alternatively, specific antibodies can be used to  
 CC screen for SOCS, which is useful as a knowledge of SOCS levels  
 CC may be important for the diagnosis of certain cancers. Soluble  
 CC SOCS polypeptides can be used to treat disease, injury or  
 CC abnormality involving cytokine mediated cellular responsiveness,  
 CC e.g. hyperimmunity, immunosuppression, allergies and hypertension.

XX Sequence 2019 BP; 427 A; 510 C; 545 G; 528 T; 9 other;



CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy.  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 4961 BP; 1411 A; 1065 C; 1123 G; 1362 T; 0 other;

Query Match 74.5%; Score 16.4; DB 21; Length 4961;  
 Best Local Similarity 94.4%; Pred. No. 48;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 agttctgtgaacatgcac 21  
 |||||  
 Db 2639 AGTTCGTGAACATGCAC 2622

RESULT 11  
 AAC12754/c  
 ID RAC12754 standard; cDNA; 226 BP.  
 XX  
 AC AAC12754;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 16829.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GSET ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PS  
 PS Claim 1; SEQ ID 16829; 71pp + CD-ROM; English.  
 XX

CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX

SQ Sequence 226 BP; 75 A; 29 C; 51 G; 67 T; 4 other;

Query Match 71.8%; Score 15.8; DB 21; Length 226;  
 Best Local Similarity 89.5%; Pred. No. 62;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 cagttctgtgaacatgcac 21  
 |||||  
 Db 138 CAGTTCGTGAATTGCAC 120

RESULT 12  
 AAH51601/c  
 ID AAH51601 standard; DNA; 319608 BP.  
 XX  
 AC AAH51601;  
 XX  
 DT 29-AUG-2001 (first entry)  
 XX  
 DE Human chromosome 13q31-q33 genomic nucleotide sequence.  
 XX

KW sbg1; q34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;  
 KW biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.  
 XX

OS Homo sapiens.

PN WO200058510-A2.

PD 05-OCT-2000.

PF 30-MAR-2000; 2000WO-IB00435.

PR 30-MAR-1999; 99US-0126903.

PR 30-APR-1999; 99US-0131971.

PR 30-APR-1999; 99US-0132065.

PR 14-JUL-1999; 99US-0143928.

PR 27-JUL-1999; 99US-0145915.

PR 29-JUL-1999; 99US-0146452.

PR 29-JUL-1999; 99US-0146453.

PR 28-OCT-1999; 99US-0162288.

XX (GSET ) GENSET.

XX Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihaï B;

PI Essioux L;

XX WPI; 2000-619082/59.

XX Polynucleotides comprising sequences from sbg1 and g35018 biallelic  
 PT markers are used for genotyping and detecting schizophrenia or bipolar  
 PT disorder and predisposition to these disorders -

PS Claim 1; Page 409-493; 737pp; English.

XX AAH51601 represents a human genomic nucleotide sequence comprising sbg1,  
 CC g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the  
 CC human chromosome 13q31-q33 locus. The nucleotide sequences contain  
 CC biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and  
 CC AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein  
 CC products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018  
 CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used  
 CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are  
 CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of  
 CC amplicons which comprise biallelic markers located on the chromosome  
 CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Biallelic markers

CC are represented in the sequences by degenerate/undefined base codes. PCR  
 CC primers AAH51818 and AAH51819 are used in the isolation of sequences of  
 CC the invention. The biallelic marker containing nucleotide sequences are  
 CC used to determine the identity of the nucleotide at a biallelic marker in  
 CC a sample DNA sequence. The nucleotide sequences may be labelled and used  
 CC for genotyping by determining the identity of a nucleotide at a region  
 CC D-related biallelic marker in a biological sample from single or multiple  
 CC subjects. By determining the frequency of a biallelic marker in a  
 CC population an association between a genotype and a trait, a haplotype and  
 CC a trait and a phenotype and a trait can be detected. The sequences can be  
 CC used to determine a predisposition to or early onset of schizophrenia or  
 CC bipolar disorder or a beneficial response to or side effects related to  
 CC treatment against schizophrenia or bipolar disorder.

XX SQ Sequence 319608 BP; 101600 A; 56677 C; 58335 G; 102722 T; 274 other;

Query Match 71.8%; Score 15.8; DB 21; Length 319608;  
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20  
 ||||| |||||  
 Db 164729 GCAGTTCTCGACATCTA 164711

## RESULT 13

AA509301/c  
 ID AA509301 standard; DNA; 319608 BP.

XX AC AA509301;

XX DT 26-SEP-2001 (first entry)

XX XX Human schizophrenia associated gene g35030 and biallelic markers A1-A71.

XX KW Human; g35030; biallelic marker; A1-A71; chromosome 13q31-q33;  
 XX KW schizophrenia; bipolar disorder; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT primer\_bind 7938..7958

FT FT /\*tag= a  
 FT /\*note= "Binds primer 99-27943.rp"

FT FT 8297..8315

FT FT /\*tag= b  
 FT /\*note= "Binds primer 99-27943-150.mis"

FT FT 8304..8328

FT FT /\*tag= c  
 FT /\*bound\_moiety= Probe\_99-27943-150

FT FT 8316

FT FT /\*tag= d  
 FT /\*note= "Biallelic marker A1"

FT FT complement (8317..8335)

FT FT primer\_bind

FT FT /\*tag= e  
 FT /\*note= "Binds primer 99-27943-150.mis complement"

FT FT complement (8446..8465)

FT FT /\*tag= f  
 FT /\*note= "Binds primer 99-27943.pu complement"

FT FT 21365..21385

FT FT /\*tag= g  
 FT /\*note= "Binds primer 99-27935.rp"

FT FT 21653..21671

FT FT /\*tag= h  
 FT /\*note= "Binds primer 99-27935-193.mis"

FT FT 21660..21684

FT FT /\*tag= i  
 FT /\*bound\_moiety= Probe\_99-27935-193

FT FT 21672

FT FT /\*tag= j  
 FT /\*note= "Biallelic marker A2"

FT FT complement (21673..21691)

FT primer\_bind

FT

FT

FT primer\_bind  
 FT /\*tag= k  
 FT /\*note= "Binds primer 99-27935-193.mis complement"  
 FT complement (21845..21864)  
 FT /\*tag= l  
 FT /\*note= "Binds primer 99-27935.pu complement"  
 FT 65463..65471  
 FT /\*tag= m  
 FT /\*note= "Binds primer 8-128.pu"  
 FT 65466..65484  
 FT /\*tag= n  
 FT /\*note= "Binds primer 8-128-33.mis"  
 FT 65473..65497  
 FT /\*tag= o  
 FT /\*bound\_moiety= Probe\_8-128-33  
 FT 65485  
 FT /\*tag= p  
 FT /\*note= "Biallelic marker A3"  
 FT complement (65486..65504)  
 FT /\*tag= q  
 FT /\*note= "Binds primer 8-128-33.mis complement"  
 FT complement (65856..65874)  
 FT /\*tag= r  
 FT /\*note= "Binds primer 8-128.rp complement"  
 FT 95034..95053  
 FT /\*tag= s  
 FT /\*note= "Binds primer 99-31960.pu"  
 FT 95377..95395  
 FT /\*tag= t  
 FT /\*note= "Binds primer 99-31960-363.mis"  
 FT 95384..95408  
 FT /\*tag= u  
 FT /\*bound\_moiety= Probe\_99-31960-363  
 FT 95396  
 FT /\*tag= v  
 FT /\*note= "Biallelic marker A4"  
 FT complement (95397..95415)  
 FT /\*tag= w  
 FT /\*note= "Binds primer 99-31960-363.mis complement"  
 FT complement (95543..95563)  
 FT /\*tag= x  
 FT /\*note= "Binds primer 99-31960.rp complement"  
 FT 107022..107040  
 FT /\*tag= y  
 FT /\*note= "Binds primer 99-24656.pu"  
 FT 107262..107280  
 FT /\*tag= z  
 FT /\*note= "Binds primer 99-24656-260.mis"  
 FT 107269..107293  
 FT /\*tag= aa  
 FT /\*bound\_moiety= Probe\_99-24656-260  
 FT 107281  
 FT /\*tag= ab  
 FT /\*note= "Biallelic marker A5"  
 FT complement (107282..107300)  
 FT /\*tag= ac  
 FT /\*note= "Binds primer 99-24656-260.mis complement"  
 FT complement (107495..107513)  
 FT /\*tag= ad  
 FT /\*note= "Binds primer 99-24656.rp complement"  
 FT 160279..160298  
 FT /\*tag= ae  
 FT /\*note= "Binds primer 99-24639.rp"  
 FT 160621..160639  
 FT /\*tag= af  
 FT /\*note= "Binds primer 99-24639-163.mis"  
 FT 160628..160652  
 FT /\*tag= ag  
 FT /\*bound\_moiety= Probe\_99-24639-163  
 FT 160640  
 FT /\*tag= ah  
 FT /\*note= "Biallelic marker A6"  
 FT complement (160641..160659)  
 FT /\*tag= ai  
 FT

```

FT primer_bind /note= "Binds primer 99-24639-163.mis complement"
FT 160770..160787 /*tag= aj
FT /note= "Binds primer 99-24634.pu"
FT complement (160785..160802)
FT /tag= ak
FT /note= "Binds primer 99-24639.pu complement"
FT 160857..160875 /*tag= al
FT /note= "Binds primer 99-24634-108.mis"
FT 160864..160888 /*tag= am
FT /bound_moiety= Probe_99-24634-108
FT 160876
FT /tag= an
FT /note= "Biallelic marker A7"
FT complement (160877..160895)
FT /tag= ao
FT /note= "Binds primer 99-24634-108.mis complement"
FT 168813..168830 /*tag= ap
FT /note= "Binds primer 99-24634-108.mis complement"
FT 168813..168830 /*tag= aq
FT /note= "Binds primer 99-7652.pu"
FT 168955..168973 /*tag= ar
FT /note= "Binds primer 99-7652-162.mis"
FT 168962..168986 /*tag= as
FT /bound_moiety= Probe_99-7652-162
FT 168974
FT /tag= at
FT /note= "Biallelic marker A8"
FT complement (168975..168993)
FT /tag= au
FT /note= "Binds primer 99-7652-162.mis complement"
FT 169331..169351 /*tag= av
FT /note= "Binds primer 99-7652.rp complement"
FT 170666..170686 /*tag= aw
FT /note= "Binds primer 99-16100.pu"
FT 170791..170809 /*tag= ax
FT /note= "Binds primer 99-16100-147.mis"
FT 170798..170822 /*tag= ay
FT /bound_moiety= Probe_99-16100-147
FT 170810
FT /tag= az
FT /note= "Biallelic marker A9"
FT complement (170811..170829)
FT /tag= ba
FT /note= "Binds primer 99-16100-147.mis complement"
FT 173346..173370 /*tag= bb
FT /note= "Binds primer 99-16100.rp complement"
FT 173065..173085 /*tag= bc
FT /note= "Binds primer 99-5862.rp"
FT 173339..173357 /*tag= bd
FT /note= "Binds primer 99-5862-167.mis"
FT 173346..173370 /*tag= be
FT /bound_moiety= Probe_99-5862-167
FT 173358
FT /tag= bf
FT /note= "Biallelic marker A10"
FT complement (173359..173377)
FT /tag= bg
FT /note= "Binds primer 99-5862-167.mis complement"

```

```

FT primer_bind complement (173495..173514)
FT /*tag= bh
FT /note= "Binds primer 99-5862.pu complement"
FT 189753..189771
FT /*tag= bi
FT /note= "Binds primer 99-5919.pu"
FT 189938..189956 /*tag= bj
FT /note= "Binds primer 99-5919-215.mis"
FT

Query Match 71.8%; Score 15.8; DB 22; Length 319608;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20
|||||
Db 164729 GCAGTCTCTGGAACATCTA 164711

RESULT 14
AAAX24719
ID AAX24719 standard; cDNA; 545 BP.
XX
AC AAX24719;
XX
DT 21-JUN-1999 (first entry)
XX
DE Tobacco leaf polyphenol oxidase cDNA clone TOBPP06.
XX
KW Polyphenol oxidase; banana; tobacco; pineapple; transgenic plant;
KW vaccine; browning; spoilage; ds.
XX
OS Nicotiana tabacum.
XX
PN WO9853080-A1.
XX
PD 26-NOV-1998.
XX
PF 19-MAY-1998; 98WO-AU00362.
XX
PR 19-MAY-1997; 97AU-0006849.
XX
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI Robinson SP;
XX
WPI; 1999-070152/06.
DR P-PSDB; AAW97990.
XX

Nucleic acid encoding polyphenol oxidase from banana, tobacco and
pineapple - useful for, e.g. increasing levels of the enzyme
expression, which is responsible for browning and spoilage of fruits
after injury or damage

Example 2; Fig 5; 47pp; English.

This partial cDNA clone, termed TOBPP06, encodes a polyphenol oxidase
(PPO) polypeptide of tobacco (see also AAW97990). The clone was
obtained by PCR amplification of young tobacco leaf cDNA using
primers (see AAX24708-12) based on conserved copper binding sites of
plant PPOs, and identified on the basis of homology to known plant
PPO genes. PPO is the major enzyme responsible for browning and
spoilage of fruits and vegetables after damage or injury. Sense
PPO nucleic acids can be used to increase the levels of PPO in a
plant. Optionally modified sense sequences and antisense sequences
can be used to reduce PPO levels (by co-suppression in the case of
sense sequences). Transgenic plants including antisense PPO
sequences, and a plant vaccine including a nucleic acid encoding
banana, tobacco or pineapple PPO or an antisense sequence are
claimed.

Sequence 545 BP; 171 A; 115 C; 115 G; 144 T; 0 other;

```

Query Match 70.9%; Score 15.6; DB 20; Length 545;  
Best Local Similarity 81.8%; Pred. No. 90;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||| ||||| | ||| |||  
Db 124 cgcggttctgtaatcatggacc 145

RESULT 15  
AAX30779  
ID AAX30779: standard; DNA; 769 BP.  
XX  
AC AAX30779;  
XX  
XX 20-MAY-1999 (first entry)  
XX  
DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:56.  
XX  
KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
KW streptococcal infection; pneumococcal; ss.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9737026-A1.  
XX  
PD 09-OCT-1997.  
XX  
PF 01-APR-1997; 97WO-US05306.  
XX  
PR 22-AUG-1996; 96US-0025788.  
PR 02-APR-1996; 96US-0014690.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
PI Stodola RK;  
XX  
DR WPI; 1997-503111/46.  
DR P-PSDB; AAY1181.  
XX  
PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
PT vaccines, drug screening, etc  
XX  
PS Claim 5; Page 91; 354pp; English.  
XX  
CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from  
CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,  
CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
CC streptococcal infections and in assays for identifying compounds that  
CC inhibit or activate the activity of the proteins. The antagonists can  
CC be used to treat an individual having need to inhibit a bacterial  
CC protein. Vectors expressing the proteins can be used to induce a  
CC protective immune response in mammals.  
XX  
SQ Sequence 769 BP; 204 A; 202 C; 140 G; 223 T; 0 other;

Query Match 70.9%; Score 15.6; DB 18; Length 769;  
Best Local Similarity 81.8%; Pred. No. 95;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||| ||||| | ||| |||  
Db 5 cgcagttctgttaccacgacc 26

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:09:23 ; Search time 353.79 Seconds  
(without alignments)  
14.083 Million cell updates/sec

Title: US-09-698-903B-14

Perfect score: 22  
Sequence: 1 cgcagttctgtgaacatgacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	100.0	3113	1 US-08-146-422-20	Sequence 20, Appl
C 2	22	100.0	3113	1 US-08-626-554-2	Sequence 2, Appl
C 3	17.2	78.2	587	1 US-08-375-241-3	Sequence 3, Appl
C 4	17.2	78.2	587	5 PCT-US92-06617A-3	Sequence 3, Appl
C 5	17.2	78.2	1174	1 US-08-375-241-5	Sequence 5, Appl
C 6	17.2	78.2	1174	5 PCT-US92-06617A-5	Sequence 5, Appl
C 7	17.2	78.2	1188	1 US-08-375-241-9	Sequence 9, Appl
C 8	17.2	78.2	1188	5 PCT-US92-06617A-9	Sequence 9, Appl
C 9	17.2	78.2	1196	1 US-08-375-241-7	Sequence 7, Appl
C 10	17.2	78.2	1196	5 PCT-US92-06617A-7	Sequence 7, Appl
C 11	15.2	69.1	195	1 US-08-158-189-9	Sequence 9, Appl
C 12	15.2	69.1	1126	1 US-08-233-788A-48	Sequence 48, Appl
C 13	15.2	69.1	1576	1 US-08-689-974-2	Sequence 2, Appl
C 14	15.2	69.1	1576	3 US-09-058-376-2	Sequence 2, Appl
C 15	15.2	69.1	2048	1 US-07-602-608-11	Sequence 11, Appl
C 16	15.2	69.1	2048	1 US-08-261-578-11	Sequence 11, Appl
C 17	15.2	69.1	2158	1 US-07-602-608-1	Sequence 1, Appl
C 18	15.2	69.1	2158	1 US-08-261-578-1	Sequence 1, Appl
C 19	14.8	67.3	30	1 US-07-602-608-18	Sequence 18, Appl
C 20	14.8	67.3	30	1 US-08-261-578-18	Sequence 18, Appl
C 21	14.8	67.3	425	4 US-09-328-111-19	Sequence 19, Appl
C 22	14.6	66.4	952	4 US-09-174-768-2	Sequence 2, Appl
C 23	14.6	66.4	1062	2 US-08-468-819-92	Sequence 92, Appl
C 24	14.6	66.4	3330	1 US-08-149-103-1	Sequence 1, Appl
C 25	14.6	66.4	3330	1 US-08-451-883-1	Sequence 1, Appl
C 26	14.6	66.4	3656	1 US-08-393-734-1	Sequence 1, Appl
C 27	14.6	66.4	3656	4 US-08-894-489-1	Sequence 1, Appl

28	14.6	66.4	4649	6	5183745-1	Patent No. 5183745
29	14.6	66.4	6443	6	5183745-5	Patent No. 5183745
C 30	14.6	66.4	9592	1	US-08-393-734-3	Sequence 3, Appl
C 31	14.6	66.4	9592	4	US-08-894-489-3	Sequence 3, Appl
C 32	14.6	66.4	80161	3	US-09-036-987A-1	Sequence 1, Appl
C 33	14.6	66.4	80161	4	US-09-370-700-1	Sequence 1, Appl
34	14.4	65.5	1929	5	PCT-US93-00031-18	Sequence 18, Appl
35	14.4	65.5	1932	5	PCT-US93-00031-20	Sequence 20, Appl
36	14.4	65.5	1941	5	PCT-US93-00031-10	Sequence 10, Appl
37	14.4	65.5	1941	5	PCT-US93-00031-22	Sequence 22, Appl
38	14.4	65.5	2205	5	PCT-US93-00031-12	Sequence 12, Appl
39	14.4	65.5	2208	5	PCT-US93-00031-14	Sequence 14, Appl
40	14.4	65.5	2217	5	PCT-US93-00031-8	Sequence 8, Appl
41	14.4	65.5	2220	5	PCT-US93-00031-16	Sequence 16, Appl
42	14.4	65.5	2811	4	US-08-482-073-3	Sequence 3, Appl
43	14.4	65.5	2813	2	US-08-344-155C-99	Sequence 99, Appl
44	14.4	65.5	2813	4	US-09-009-490A-90	Sequence 90, Appl
45	14.4	65.5	3080	4	US-08-482-073-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-08-146-422-20/c  
; Sequence 20, Application US/08146422  
; Patent No. 5543576  
; GENERAL INFORMATION:  
; APPLICANT: VAN OOLJEN, ALBERT J. J.  
; APPLICANT: RIETVELD, KRIJN  
; APPLICANT: HOEKEMA, ANDREAS  
; APPLICANT: PEN, JAN  
; APPLICANT: SIJMONS, PETER C.  
; APPLICANT: VERWOERD, TEUNIS C.  
; APPLICANT: QUAX, WILHEMUS J.  
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
; TITLE OF INVENTION: USE  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/146,422  
; FILING DATE: 02-NOV-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KENNEDY, BILL  
; REGISTRATION NUMBER: 33,407  
; REFERENCE/DOCKET NUMBER: 44615-20011.23  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3113 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-146-422-20

Query Match 100.0%; Score 22; DB 1; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.012;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 1183 CGCAGTTCTGTGAACATCGACC 1162

## RESULT 2

US-08-626-554-2/c  
 ; Sequence 2, Application US/08626554  
 ; Patent No. 5714474  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN OOLJEN, ALBERT J.J.  
 ; APPLICANT: RIETVELD, KRJON  
 ; APPLICANT: HOEKEMA, ANDREAS  
 ; APPLICANT: PEN, JAN  
 ; APPLICANT: SIJMONS, PETER C.  
 ; APPLICANT: VERWOERD, TENNIS C.  
 ; APPLICANT: QUAX, WILHEMUS J.  
 ; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
 ; TITLE OF INVENTION: USE  
 ; NUMBER OF SEQUENCES: 32  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 PENNSYLVANIA AVENUE NW  
 ; CITY: WASHINGTON  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1888  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/626,554  
 ; FILING DATE: 02-APR-1996  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 26192-20011.10  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030 MRSNFOERSMWSH  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3113 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-626-554-2

Query Match 100.0%; Score 22; DB 1; Length 3113;  
 Best Local Similarity 100.0%; Pred. No. 0.012;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 1183 CGCAGTTCTGTGAACATCGACC 1162

## RESULT 3

US-08-375-241-3  
 ; Sequence 3, Application US/08375241  
 ; Patent No. 5648481  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Parodos, Kyriaki  
 ; APPLICANT: McCarty, Janice  
 ; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
 ; TITLE OF INVENTION: Shigella

; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02173

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/375,241  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/738,800  
 ; FILING DATE: 31-JUL-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Granahan, Patricia  
 ; REGISTRATION NUMBER: 32,227  
 ; REFERENCE/DOCKET NUMBER: GTR90-04  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9540  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 587 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-375-241-3

Query Match 78.2%; Score 17.2; DB 1; Length 587;  
 Best Local Similarity 86.4%; Pred. No. 3.1;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 566 CGCAGTACTGTGAACCTCGATC 587

## RESULT 4

PCT-US92-06617A-3  
 ; Sequence 3, Application PC/TUS9206617A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Parodos, Kyriaki  
 ; APPLICANT: McCarty, Janice  
 ; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
 ; TITLE OF INVENTION: Shigella  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Amoco Corporation  
 ; STREET: 200 East Randolph Drive, P.O. Box 87703  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: U.S.A.  
 ; ZIP: 60680  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/06617A  
 ; FILING DATE: 19920728  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/738,800  
 ; FILING DATE: 31-JUL-1991  
 ; ATTORNEY/AGENT INFORMATION:

; NAME: Galloway, Norval B.  
 ; REGISTRATION NUMBER: 33,595  
 ; REFERENCE/DOCKET NUMBER: GTR90-04 PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-856-7180  
 ; TELEFAX: 312-856-4972  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 587 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ;  
 PCT-US92-06617A-3

Query Match 78.2%; Score 17.2; DB 5; Length 587;  
 Best Local Similarity 86.4%; Pred. No. 3.1;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 566 CGCAGTACTGTGAACCTCGATC 587  
 ||||| ||||| ||||| |||||

RESULT 5  
 US-08-375-241-5  
 ; Sequence 5, Application US/08375241  
 ; Patent No. 5648481  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Parodos, Kyriaki  
 ; APPLICANT: McCarty, Janice  
 ; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
 ; TITLE OF INVENTION: Shigella  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/375,241  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/738,800  
 ; FILING DATE: 31-JUL-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Granahan, Patricia  
 ; REGISTRATION NUMBER: 32,227  
 ; REFERENCE/DOCKET NUMBER: GTR90-04  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9540  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1174 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ;  
 us-08-375-241-5

Query Match 78.2%; Score 17.2; DB 1; Length 1174;  
 Best Local Similarity 86.4%; Pred. No. 3.6;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 322 CGCAGTACTGTGAACCTCGATC 343  
 ||||| ||||| ||||| |||||

RESULT 6  
 PCT-US92-06617A-5  
 ; Sequence 5, Application PC/TUS9206617A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Parodos, Kyriaki  
 ; APPLICANT: McCarty, Janice  
 ; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
 ; TITLE OF INVENTION: Shigella  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Amoco Corporation  
 ; STREET: 200 East Randolph Drive, P.O. Box 87703  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: U.S.A.  
 ; ZIP: 60680  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/06617A  
 ; FILING DATE: 19920728  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/738,800  
 ; FILING DATE: 31-JUL-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Galloway, Norval B.  
 ; REGISTRATION NUMBER: 33,595  
 ; REFERENCE/DOCKET NUMBER: GTR90-04 PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-856-7180  
 ; TELEFAX: 312-856-4972  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1174 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ;  
 PCT-US92-06617A-5

Query Match 78.2%; Score 17.2; DB 5; Length 1174;  
 Best Local Similarity 86.4%; Pred. No. 3.6;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 Db 322 CGCAGTACTGTGAACCTCGATC 343  
 ||||| ||||| ||||| |||||

RESULT 7  
 US-08-375-241-9  
 ; Sequence 9, Application US/08375241  
 ; Patent No. 5648481  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Parodos, Kyriaki  
 ; APPLICANT: McCarty, Janice  
 ; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
 ; TITLE OF INVENTION: Shigella  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ;  
 us-08-375-241-9

```
;
;
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,241
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: GTR90-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1188 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-375-241-9

Query Match 78.2%; Score 17.2; DB 1; Length 1188;
Best Local Similarity 86.4%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 8
US-08-375-241-7
; Sequence 9, Application PC/US9206617A
; Patent No. 5648481
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyriaki
; APPLICANT: McCarthy, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corporation
; STREET: 200 East Randolph Drive, P.O. Box 87703
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06617A
; FILING DATE: 19920728
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Calloway, Norval B.
; REGISTRATION NUMBER: 33,595
; REFERENCE/DOCKET NUMBER: GTR90-04 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-856-7180
; TELEFAX: 312-856-4972
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1188 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US92-06617A-9
```

```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1188 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US92-06617A-9

Query Match 78.2%; Score 17.2; DB 5; Length 1188;
Best Local Similarity 86.4%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 9
US-08-375-241-7
; Sequence 7, Application US/08375241
; Patent No. 5648481
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyriaki
; APPLICANT: McCarthy, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,241
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: GTR90-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-375-241-7

Query Match 78.2%; Score 17.2; DB 1; Length 1196;
Best Local Similarity 86.4%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 10
PCT-US92-06617A-7
```

```
; Sequence 7, Application PC/TUS9206617A
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyrilaki
; APPLICANT: McCarty, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corporation
; STREET: 200 East Randolph Drive, P.O. Box 87703
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION NUMBER: PCT/US92/06617A
; FILING DATE: 19920728
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, Norval B.
; REGISTRATION NUMBER: 33,595
; REFERENCE/DOCKET NUMBER: GTR90-04 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-856-7180
; TELEFAX: 312-856-4972
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US92-06617A-7

Query Match 78.2%; Score 17.2; DB 5; Length 1196;
Best Local Similarity 86.4%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
      ||||| ||||| ||||| |||||
Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 11
US-08-158-189-9
; Sequence 9, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/158,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,232
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnson, Philip S.
; REGISTRATION NUMBER: 27,200
; REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-158-189-9

Query Match 69.1%; Score 15.2; DB 1; Length 195;
Best Local Similarity 85.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcga 20
      ||||| ||||| ||||| |||||
Db 130 CACAGTTCAGTGAGCATCGA 149

RESULT 12
US-08-233-788A-48/c
; Sequence 48, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1126 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(3..23, 27..944, 948..1124)
US-08-233-788A-48
```

```
Query Match          69.1%; Score 15.2; DB 1; Length 1126;
Best Local Similarity 85.0%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 gcagttctgtgaacatcgacc 21
   ||||| ||||| ||||| |||||
Db 492 GCTGTCGGTGAACATCGCC 473
```

```
RESULT 13
US-08-689-974-2/c
; Sequence 2, Application US/08689974
; Patent No. 5776732
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Murray, Lynn E.
; TITLE OF INVENTION: NOVEL HUMAN INDUCED TUMOR PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,974
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0113 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: BRAIN0T03
; CLONE: 530522
```

```
US-08-689-974-2
Query Match          69.1%; Score 15.2; DB 1; Length 1576;
Best Local Similarity 85.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 cagttctgtgaacatcgacc 22
   ||||| ||||| ||||| |||||
Db 893 CAGCTCGGTGAACATCGTCC 874
```

```
RESULT 14
US-09-058-376-2/c
; Sequence 2, Application US/09058376
```

```
; Patent No. 6080841
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Murray, Lynn E.
; TITLE OF INVENTION: NOVEL HUMAN INDUCED TUMOR PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,376
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,974
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0113 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: BRAIN0T03
; CLONE: 530522
; US-09-058-376-2
```

```
Query Match          69.1%; Score 15.2; DB 3; Length 1576;
Best Local Similarity 85.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 cagttctgtgaacatcgacc 22
   ||||| ||||| ||||| |||||
Db 893 CAGCTCGGTGAACATCGTCC 874
```

```
RESULT 15
US-07-602-608-11/c
; Sequence 11, Application US/07602608
; Patent No. 5382524
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/602,608
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-07-602-608-11

```

```

Query Match          69.1%; Score 15.2; DB 1; Length 2048;
Best Local Similarity 85.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 cgcagttctgtgaacatcga 20
    ||||| ||||| |||||
Db 101 CGCAGGTCGTGGACATCAA 82

```

Search completed: February 15, 2002, 19:09:24  
Job time: 14738 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:04:22 ; Search time 9904.61 Seconds  
(without alignments)  
23.868 Million cell updates/sec

Title: US-09-698-903B-14  
Perfect score: 22  
Sequence: 1 cgcagttctggaacatgcacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	17.8	80.9	653	10	BE377681	BE377681 601229824
2	16.8	76.4	244	10	AI561687	AI561687 vw88f09.x
3	16.8	76.4	279	13	AZ777264	AZ777264 2M0011K06
4	16.8	76.4	346	10	AI019782	AI019782 ua90h12.r
5	16.8	76.4	605	13	FR0030063	AL026432 Fugu rubr
6	16.8	76.4	611	13	FR0030058	AL026427 Fugu rubr
7	16.8	76.4	613	13	AZ777336	AZ777336 2M0011I08
8	16.8	76.4	1301	12	AK006515	AK006515 Mus muscu
9	16.4	74.5	206	10	AW086177	AW086177 xc77e11.x
10	16.4	74.5	234	10	AA701011	AA701011 zg55d04.s
11	16.4	74.5	270	11	F02043	F02043 HSCOMG102 n
12	16.4	74.5	290	10	AA569516	AA569516 nf23e01.s

13	16.4	74.5	303	11	F03921	F03921 HSC2DF112 n
14	16.4	74.5	304	10	AA843570	AA843570 aj54f09.s
15	16.4	74.5	335	10	AI084952	AI084952 ow86c05.s
16	16.4	74.5	399	10	AI658711	AI658711 tu22f10.x
17	16.4	74.5	412	11	BF066458	BF066458 st15b08.y
18	16.4	74.5	435	11	N49244	N49244 yy83f09.sl
19	16.4	74.5	441	10	AW196345	AW196345 xm32al2.x
20	16.4	74.5	455	11	BF483372	BF483372 WHEI1794.A
21	16.4	74.5	460	10	AA679411	AA679411 z129b11.s
22	16.4	74.5	463	10	AA426087	AA426087 zv52c11.s
23	16.4	74.5	483	10	AI149964	AI149964 qf38g01.x
24	16.4	74.5	487	10	AI802194	AI802194 tx25c03.x
25	16.4	74.5	499	10	AA421463	AA421463 zu06s09.s
26	16.4	74.5	508	10	AA700220	AA700220 z344h06.s
27	16.4	74.5	526	10	AW235369	AW235369 xm56b05.x
28	16.4	74.5	545	11	R60174	R60174 yhl12f10.sl
29	16.4	74.5	553	10	AI631483	AI631483 wa89e08.x
30	16.4	74.5	560	10	AW013257	AW013257 Sp173f.wl
31	16.4	74.5	582	11	W91993	W91993 zh47d10.sl
32	16.4	74.5	612	13	BH121721	BH121721 RPCI-24-2
33	16.4	74.5	630	10	AA700001	AA700001 z169b07.s
34	16.4	74.5	650	10	AW952652	AW952652 EST364737
35	16.4	74.5	673	10	AA868429	AA868429 ak42b04.s
36	16.4	74.5	679	10	AI680857	AI680857 tx42b01.x
37	16.4	74.5	706	13	AZ717353	AZ717353 RPCI-24-1
38	16.4	74.5	758	10	AI740628	AI740628 wq23f08.x
39	16.4	74.5	768	10	AI637690	AI637690 tt29e11.x
40	16.4	74.5	804	10	AL041092	AL041092 DXF2p434C
41	16.4	74.5	833	10	BE412370	BE412370 JTL005.F0
42	16.2	73.6	122	11	BG989620	BG989620 PM2-HT117
43	16.2	73.6	400	13	AQ852470	AQ852470 LMAJFV1.1
44	16.2	73.6	406	13	AZ160523	AZ160523 SP_0066_A
45	16.2	73.6	431	10	AV655742	AV655742 AV655742

#### ALIGNMENTS

RESULT 1

LOCUS BE377681 653 bp mRNA EST 21-JUL-2000  
DEFINITION 601229824F1 NCI\_CGAP\_Mam1 Mus musculus cdna clone IMAGE:3594076 5',  
mRNA sequence.

ACCESSION BE377681

VERSION BE377681.1 GI:9323046

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 653)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM8767 row: o column: 05

High quality sequence stop: 612.

Location/Qualifiers

1. .653

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:3594076"

/clone\_lib="NCI\_CGAP\_Mam1"

/tissue\_type="tumor, biopsy sample"

```
/dev_stage="3 months, virgin"
```

```
/lab_host="DH108"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
```

```
BASE COUNT
ORIGIN
```

```
132 a 181 c 195 g 145 t
```

## Query Match

```
Best Local Similarity 80.9%; Score 17.8; DB 10; Length 653;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY

```
1 cgcagttctgtgaacatgcac 21
```

DB

```
193 CGCAGTCTGTGAGCTGCAC 213
```

## RESULT 2

```
AI561687 244 bp mRNA EST 25-MAR-1999
LOCUS vw88f09.x1 StrataGene mouse skin (#937313) Mus musculus cDNA clone
DEFINITION IMAGE:1262057 3', mRNA sequence.
ACCESSION AI561687
VERSION AI561687
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
```

```
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 244)
Marra, M., Hallier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person,
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R., and Wilson, R.
```

```
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
```

```
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
```

```
MGI:664609
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
```

```
High quality sequence stop: 229.
Location/Qualifiers
1. .244
```

```
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1262057"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
```

```
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGCTTTTTTTTTTTTTTT 3"
```

```
53 a 53 c 51 g 87 t
```

## BASE COUNT

```
ORIGIN
```

```
76.4%; Score 16.8; DB 10; Length 244;
```

```
Query Match
```

```
Best Local Similarity 90.0%; Pred. No. 5.3e+02;
```

```
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 3 cagttctgtgaacatgcac 22
```

```
DB 85 CAGTCTCTTGAACATCTACC 104
```

```
RESULT 3
AZ777264 279 bp DNA GSS 16-FEB-2001
LOCUS 2M0011K06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0011K06 R, DNA sequence.
ACCESSION AZ777264
VERSION AZ777264
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
```

```
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 279)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,
and Wright, D., Weiss, R.
```

```
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0011 row: K column: 06
Seq primer: CACACAGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 279.
Location/Qualifiers
1. .279
```

```
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0011K06"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

```
80 a 65 c 69 g 65 t
```

```
BASE COUNT
ORIGIN
```

Query Match 76.4%; Score 16.8; DB 13; Length 279;  
Best Local Similarity 90.0%; Pred. No. 5.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcga 20  
|||||  
Db 162 CTCAGTTCTGTGAATATCGA 181

RESULT 4  
LOCUS AI019782 346 bp mRNA EST 16-JUN-1998  
DEFINITION ua90h12.r1 Soares\_mammary\_gland\_NBMWG Mus musculus cDNA clone  
IMAGE:1364807 5', mRNA sequence.  
ACCESSION AI019782  
VERSION AI019782.1 GI:3234118  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 346)  
REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:898027  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 285.

FEATURES  
source  
1..346  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1364807"  
/clone\_lib="Soares\_mammary\_gland\_NBMWG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/notes="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia  
) with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGGAATGGTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT 74 a 95 c 120 g 57 t  
ORIGIN  
|||||

Query Match 76.4%; Score 16.8; DB 10; Length 346;  
Best Local Similarity 90.0%; Pred. No. 5.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ggcagttctgtgaacatcga 21  
|||||

Db 182 GCAGTTCTGTGACCATCCAC 201

RESULT 5  
LOCUS FR0030063 605 bp DNA GSS 25-JUN-1998  
DEFINITION Fugu rubripes GSS sequence, clone 072H16aE4, genomic survey  
sequence.  
AL026432  
VERSION AL026432.1 GI:3263775  
KEYWORDS GSS; genome survey sequence.  
SOURCE Takifugu rubripes.  
ORGANISM Takifugu rubripes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Takifugu.  
1 (bases 1 to 605)  
REFERENCE Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,  
Williams,G. and Brenner,S.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource  
Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:  
biohelp@hmp.mrc.ac.uk  
COMMENT Vector: pBluescript II KS  
V\_type: phagemid  
PRIMER: KS  
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.

FEATURES  
source  
1..605  
/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid 072H16"  
/clone="072H16aE4"  
BASE COUNT 139 a 147 c 173 g 132 t 14 others  
ORIGIN  
|||||

Query Match 76.4%; Score 16.8; DB 13; Length 605;  
Best Local Similarity 90.0%; Pred. No. 6.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcga 20  
|||||  
Db 29 CCCAGTTTGTGAACATCGA 48

RESULT 6  
LOCUS FR0030058 611 bp DNA GSS 25-JUN-1998  
DEFINITION Fugu rubripes GSS sequence, clone 072H16aG6, genomic survey  
sequence.  
AL026427  
VERSION AL026427.1 GI:3263770  
KEYWORDS GSS; genome survey sequence.  
SOURCE Takifugu rubripes.  
ORGANISM Takifugu rubripes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Takifugu.  
1 (bases 1 to 611)  
REFERENCE Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,  
Williams,G. and Brenner,S.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource  
Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:  
biohelp@hmp.mrc.ac.uk  
COMMENT Vector: pBluescript II KS  
V\_type: phagemid  
PRIMER: KS





This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham.

# FEATURES

```

source
1. .234
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="CDB:1303157"
/db_xref="taxon:9606"
/clone="IMAGE:397255"
/clone_lib="Soares.pineal_gland_N3HPC"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: pineal gland; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTACCAATCTGAAGTGGCGCGCTTTTGTGTTTGTGTTT 3']
, double-stranded cDNA was size selected, ligated to Eco
RI adapters (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Donaldo."
BASE COUNT      79 a      31 c      35 g      89 t
ORIGIN

Query Match      74.5%; Score 16.4; DB 10; Length 234;
Best Local Similarity 94.4%; Pred. No. 8.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21
|||||
Db 21 AGTTCTGTGAACATTGAC 38

RESULT 11
F02043
LOCUS
DEFINITION
HSCOMG102 normalized infant brain cDNA Homo sapiens cDNA clone
c-Omg10 3', mRNA sequence.
ACCESSION F02043
VERSION F02043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 270)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read, removed at sequence 5'end
Genexpress_library_idt: C; Genexpress_sequence_idt: alc-Omg10
Seq primer: (-21)M13-universal.
Location/Qualifiers
1. .270
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-Omg10"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"

```

/note="Organ: brain; Vector: lafmid BA; Site\_1: HindIII;  
Site\_2: NotI; sex=Female; dev stage=3 months old;  
isolate=muscular atrophy patient; tissue\_type=total brain  
; total mRNA was oligo(dT) primed and directionally  
cloned 5' -> 3' into the HindIII -> NotI sites of the  
lafmid BA vector. Clone library from B.Souares, Psychiatry  
Dept. Columbia University, USA. Normalization\_method:  
Bento Soares, P.N.A.S in press"

BASE COUNT 86 a 38 c 39 g 107 t  
ORIGIN

Query Match 74.5%; Score 16.4; DB 11; Length 270;  
Best Local Similarity 94.4%; Pred. No. 8.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21  
|||||

Db 23 AGTTCTGTGAACATTGAC 40

RESULT 12  
AA569516  
LOCUS

DEFINITION  
nf23e01.s1 NCI\_CGAP\_Prl Homo sapiens cDNA clone IMAGE:914616, mRNA  
sequence.

ACCESSION AA569516  
VERSION AA569516.1 GI:2343496  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 290)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaquil, M.D.  
, Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 263.  
Location/Qualifiers  
1. .290

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:914616"  
/clone\_lib="NCI\_CGAP\_Prl"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"

/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from 5,000-10  
,000 microdissected, histologically normal prostate  
epithelial cells. Double-stranded cDNA was ligated to  
EcoRI adaptors. 5 cycles of PCR applied to the cDNA with  
an adaptor-specific primer, and the resulting PCR product  
subcloned into pAMP10 by the UDG-cloning method (Life  
technologies). Average insert size is 600 bp. NOTE: Not  
directionally cloned. This library was constructed by  
David Krizman."

BASE COUNT 77 a 76 c 73 g 64 t  
ORIGIN

FEATURES  
source

```

Query Match      74.5%; Score 16.4; DB 10; Length 290;
Best Local Similarity 94.4%; Pred. No. 8.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 gttctgtgaacatgcacc 22
||||| |||||||
Db 233 GTTCTGTGGACATCGACC 250

RESULT 13
F03921
LOCUS      F03921      303 bp      mRNA      EST      19-FEB-1995
DEFINITION HSC-2DF112 normalized infant brain cDNA Homo sapiens cDNA clone
C-2df11 3', mRNA sequence.
ACCESSION  F03921
VERSION    F03921
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 303)
AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
,M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabatchis,C. and Tessier,A.
TITLE     IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL   C.R.Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE   95277534
COMMENT   Contact: Genethon
            Genethon-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read. removed at sequence 5'end
            Genexpress_library_id: C; Genexpress_sequence_id: alc-2df11
            Seq primer: (-21)M13_universal.
FEATURES   source
            Location/Qualifiers
            1..303
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="c-2df11"
                /clone_lib="normalized infant brain cDNA"
                /sex="Female"
                /tissue_type="total brain"
                /dev_stage="3 months old"
                /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                Site_2: NotI; sex=Female; dev_stage=3 months old;
                isolate=muscular atrophy patient; tissue_type=total brain
                ; total mRNA was oligo-(dT) primed and directionally
                cloned 5' -> 3' into the HindIII -> NotI sites of the
                lafmid BA vector. Clone library from B.Souares, Psychiatry
                Dept. Columbia University, USA. Normalization_method:
                Bento Soares, P.N.A.S in press"
BASE COUNT 99 a 42 c 43 g 116 t 3 others
ORIGIN

Query Match      74.5%; Score 16.4; DB 11; Length 303;
Best Local Similarity 94.4%; Pred. No. 8.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatgcacg 21
||||| ||||||| |||
Db 23 AGTTCTGTGAACATGTGAC 40

RESULT 14
AA843570
LOCUS      AJ544952      304 bp      mRNA      EST      31-DEC-1998
DEFINITION aJ544f09.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394153
3', mRNA sequence.
ACCESSION  AA843570
VERSION    AA843570.1 GI:2930088
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 304)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrip/image/image.html
            Insert Length: 1108 Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 284.
FEATURES   source
            Location/Qualifiers
            1..304
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:1394153"
                /clone_lib="Soares_testis_NHT"
                /sex="male"
                /lab_host="DH10B"
                /note="Vector: pT73D-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                was prepared from mRNA obtained from Clontech Laboratories
                , Inc., and primed with a Not I - oligo(dT) primer [5',
                TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
                Double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT73 vector. Library
                went through one round of normalization to Cot5, and was
                constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 99 a 43 c 43 g 119 t
ORIGIN

Query Match      74.5%; Score 16.4; DB 10; Length 304;
Best Local Similarity 94.4%; Pred. No. 8.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatgcacg 21
||||| ||||||| |||
Db 26 AGTTCTGTGAACATGTGAC 43

RESULT 15
AI084952
LOCUS      AI084952      335 bp      mRNA      EST      17-AUG-1998
DEFINITION ow86c05.s1 Soares_fetal_liver_spleen_LNFLS_S1 Homo sapiens cDNA
clone IMAGE:1653704 3', mRNA sequence.
ACCESSION  AI084952
VERSION    AI084952.1 GI:3423375
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 335)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            TITLE

```

Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 152.  
 Location/Qualifiers  
 1..335  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1653704"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
 with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 This is a subtracted version of the original Soares fetal  
 liver spleen INFLS library. 1st strand cDNA was primed  
 with a Pac I - oligo(dT) primer [5',  
 AACTGGAAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Pac I and cloned into the Pac I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M.Fatima Bonaldo."  
 112 a 50 c 47 g 126 t  
 BASE COUNT  
 ORIGIN  
 Query Match 74.5%; Score 16.4; DB 10; Length 335;  
 Best Local Similarity 94.4%; Pred. No. 8.8e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 4 agttctgtgaacatcgac 21  
 |||||  
 Db 27 AGTTCTGTGAACATTGAC 44

Search completed: February 15, 2002, 18:04:26  
 Job time: 20975 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:01:19 ; Search time 868.33 Seconds  
(without alignments)  
409.741 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagtttggttcacga.....cagctggtacattgcgtag 415

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	415	100.0	415	22	Right (5') border
2	380.8	91.8	416	22	Left (3') border f
3	191.4	46.1	5864	17	Plasmid pTCS113 T-
4	191.4	46.1	5865	22	Chimeric T-DNA of
5	191.4	46.1	7599	22	Nucleotide sequenc
6	188.8	45.5	5228	22	Plasmid pTCS172del
7	188.8	45.5	6539	21	E. coli plasmid pT
8	188.8	45.5	6548	17	Plasmid pTCS174 use
9	188.8	45.5	6548	18	Plasmid pTCS172. C
10	188.8	45.5	6548	21	E. coli plasmid pT
11	188.8	45.5	7492	22	Plasmid pTCS346. U

C 12	188.4	45.4	1303	17	AAT39337	Plasmid pTCS88 (Eco
C 13	188.4	45.4	3201	12	AAQ14529	pS029 Bt ICP codi
C 14	188.4	45.4	4832	22	AAH25423	Nucleotide sequenc
C 15	188.4	45.4	4946	18	AAT59531	T-DNA of plasmid p
C 16	188.4	45.4	4946	22	AAH25422	Nucleotide sequenc
C 17	188.4	45.4	5349	19	AAV23239	T-DNA of pTCS24.
C 18	188.4	45.4	5864	17	AAT39339	Plasmid pTCS113 T-
C 19	188.4	45.4	5865	22	AAQ06990	Chimeric T-DNA of
C 20	188.4	45.4	7566	14	AAQ42160	Plasmid pS0212 co
C 21	188.4	45.4	7639	14	AAQ42159	Plasmid pJ0884 con
C 22	182.4	44.0	1037	11	AAQ04705	USP-Promoter-casse
C 23	182.4	44.0	1085	11	AAQ04703	Legumin-signalpept
C 24	182.4	44.0	1160	11	AAQ04706	USP-signalpeptide
C 25	179.4	43.2	1077	22	AAH25439	Right flanking reg
C 26	177	42.7	3201	12	AAQ15144	pVE36 Bt ICP codin
C 27	153	36.9	1186	13	AAQ25707	Chimeric neo gene
C 28	146	35.2	3153	21	AAQ29122	Plasmid pJ0884 con
C 29	146	35.2	3336	21	AAQ29121	Plasmid pJ0884 con
C 30	146	35.2	3694	21	AAQ29124	Plasmid pJ0884 con
C 31	146	35.2	3877	21	AAQ29123	Plasmid pJ0884 con
C 32	146	35.2	24593	6	AAQ50226	Sequence of opine
C 33	146	35.2	24596	6	AAQ50182	Complete nucleotide
C 34	107.6	25.9	936	22	AAF58252	Oligonucleotide D1
C 35	107.6	25.9	936	22	AAF58254	Oligonucleotide D1
C 36	107.6	25.9	936	22	AAF58257	Oligonucleotide D1
C 37	107.6	25.9	936	22	AAF58259	Oligonucleotide D2
C 38	107.6	25.9	936	22	AAF58262	Oligonucleotide D2
C 39	107.6	25.9	938	22	AAF58255	Oligonucleotide D1
C 40	106.4	25.6	936	22	AAF58252	Oligonucleotide D1
C 41	106.4	25.6	936	22	AAF58254	Oligonucleotide D1
C 42	106.4	25.6	936	22	AAF58257	Oligonucleotide D1
C 43	106.4	25.6	936	22	AAF58259	Oligonucleotide D2
C 44	106.4	25.6	936	22	AAF58262	Oligonucleotide D2
C 45	106.4	25.6	938	22	AAF58255	Oligonucleotide D1

ALIGNMENTS

RESULT 1	
AAD06997	AAD06997 standard; DNA; 415 BP.
ID	AAD06997 standard; DNA; 415 BP.
XX	
AC	AAD06997;
XX	
XX	
DT	06-AUG-2001 (first entry)
XX	
DE	Right (5') border flanking region of elite event MS-B2.
XX	
KW	MS-B2 elite event; transgenic Brassica plant; transformation event;
KW	male-sterility gene; ds.
XX	
OS	Chimeric - Agrobacterium sp.
OS	Chimeric - Brassica sp.
XX	
FT	Key Location/Qualifiers
FT	misc_feature 1..234
FT	/tag= a
FT	/note= "Corresponds to plant DNA"
FT	misc_feature 235..415
FT	/tag= b
FT	/note= "Corresponds to T-DNA"
XX	
PN	WO200131042-A2.
XX	
PD	03-MAY-2001.
XX	
PF	26-OCT-2000; 2000WO-EP10680.
XX	
PR	29-OCT-1999; 99US-0430497.
XX	
PA	(AVET ) AVENTIS CROPS SCIENCE NV.
XX	

PI Weston B, De Beuckeleer M;  
XX WPI; 2001-300517/31.  
XX  
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX  
XX Claim 11; Page 51; 53pp; English.  
XX  
XX The present invention relates to a transgenic Brassica plant or its  
CC seed, cells or tissues, characterised by harbouring a specific  
CC transformation event, particularly by the presence of a male-sterility  
CC gene, at a specific location in the Brassica genome. Transgenic  
CC Brassica plant is useful for producing a hybrid seed by crossing the  
CC transgenic plant with a male-fertile Brassica plant and harvesting the  
CC hybrid seed from the transgenic Brassica plant.  
CC The present sequence is right (5') border flanking region of elite event  
CC MS-B2.  
XX  
SQ Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 100.0%; Score 415; DB 22; Length 415;  
Best Local Similarity 100.0%; Pred. No. 3.5e-76;  
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcgagttggttgcattgatttgggtttgactcttcaccattacattgaaactct 60  
DB |||||||  
DB 1 gtcgagttggttgcattgatttgggtttgactcttcaccattacattgaaactct 60  
QY 61 taccggtgagaacaactcacaagcattatcatgttcataataatgtacattatc 120  
DB |||||||  
DB 61 taccggtgagaacaactcacaagcattatcatgttcataataatgtacattatc 120  
QY 121 gtatatatcacgtatataaaatagtagcagaagaatccatgtaaacgagggggcacc 180  
DB |||||||  
DB 121 gtatatatcacgtatataaaatagtagcagaagaatccatgtaaacgagggggcacc 180  
QY 181 atggtttcaagtattataataattataattattggttaggtgacatggtccgataa 240  
DB |||||||  
DB 181 atggtttcaagtattataataattataattattggttaggtgacatggtccgataa 240  
QY 241 gaaaaggcaattgtgagtgtaattcccatcttgaaagaataatgtttaaatattat 300  
DB |||||||  
DB 241 gaaaaggcaattgtgagtgtaattcccatcttgaaagaataatgtttaaatattat 300  
QY 301 tgataaaataaacaagtcagggtattatagcccaagcaaaaacataaattttgatgcaag 360  
DB |||||||  
DB 301 tgataaaataaacaagtcagggtattatagcccaagcaaaaacataaattttgatgcaag 360  
QY 361 tttaaattcagaataattttcaataactgatttatcagctggtgacattgccgtag 415  
DB |||||||  
DB 361 tttaaattcagaataattttcaataactgatttatcagctggtgacattgccgtag 415

RESULT 2  
AAD06999/c  
ID AAD06999 standard; DNA; 416 BP.  
XX  
AC AAD06999;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE Left (3') border flanking region of elite event MS-B2.  
XX  
KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; ds.  
XX  
OS Chimeric - Agrobacterium sp.  
OS Chimeric - Brassica sp.  
XX  
FH Key Location/Qualifiers

FT misc\_feature 1..193  
FT /\*tag= a  
FT /note= "Corresponds to T-DNA"  
FT 194..416  
FT /\*tag= b  
FT /note= "Corresponds to plant DNA"  
XX  
PN WO200131042-A2.  
XX  
XX 03-MAY-2001.  
XX  
XX 26-OCT-2000; 2000WO-EP10680.  
XX  
XX 29-OCT-1999; 99US-0430497.  
XX  
XX (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
XX Weston B, De Beuckeleer M;  
PI  
XX WPI; 2001-300517/31.  
XX  
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX  
XX Claim 11; Page 52; 53pp; English.  
XX  
XX The present invention relates to a transgenic Brassica plant or its  
CC seed, cells or tissues, characterised by harbouring a specific  
CC transformation event, particularly by the presence of a male-sterility  
CC gene, at a specific location in the Brassica genome. Transgenic  
CC Brassica plant is useful for producing a hybrid seed by crossing the  
CC transgenic plant with a male-fertile Brassica plant and harvesting the  
CC hybrid seed from the transgenic Brassica plant.  
CC The present sequence is left (3') border flanking region of elite event  
CC MS-B2.  
XX  
SQ Sequence 416 BP; 137 A; 72 C; 54 G; 152 T; 1 other;

Query Match 91.8%; Score 380.8; DB 22; Length 416;  
Best Local Similarity 98.5%; Pred. No. 3.2e-69;  
Matches 404; Conservative 1; Mismatches 3; Indels 2; Gaps 2;

QY 7 ttggtgttcattgatttgggtttgactcttcaccattacattgaaactcttcagg 66  
DB |||||||  
DB 409 TTGGGTGTTTCATGATTTTGGGTTTTCACCTTCACCATTACATTGAAACTCTTACGGA 350  
QY 67 tgagaacaactcacaagcattatcatgtttcataataataatgtacattatcac 125  
DB |||||||  
DB 349 TGAGAACAACTCACAAGCATTAATCATGTTTCAATAAATATATGTACATTATACGGTATA 290  
QY 126 tatacagctatacaaatagtagcagaagaatccatgtaaacgagggggccacatggt 185  
DB |||||||  
DB 289 TATACAGCTATACAAATAGTAGCGGAGAAATCCATGCTAAAGCAGCAGGGGCCACCTGTT 230  
QY 186 ttcaagtattataataattataattattgtaggtgacatggtccgataaagaaa 245  
DB |||||||  
DB 229 TTC-AGTATTATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 171  
QY 246 gccaattgttagtggttaattcccatcttgaagaataatagtttaataattatttgata 305  
DB |||||||  
DB 170 GCAATTTGTAGATGTTAATTTCCATCTTCAAGAAATATAGTTTAATTAATTAATTAATTA 111  
QY 306 aaataaacaagtcagggtattatagtcacaagcaaaaacataaattattgtatgcaagtttaa 365  
DB |||||||  
DB 110 AAATAACAAGTCAGGTATTATAGTCCAAAGCAAAACATAAATTTATTGTGCAAGTTAA 51  
QY 366 attcagaataatttcaataactgatttatcagctggtgacattgccgtag 415  
DB |||||||  
DB 50 ATTGAGAAATATTTCAATAACTGATTATATACAGCTGGTACATTGCCGCTAG 1



```

FT      misc_feature      1..25
FT      /tag= a
FT      /note= "Right border repeat from the TL-DNA from
FT      pTiB6S3"
FT      26..53
FT      misc_feature
FT      /tag= b
FT      /note= "Synthetic polylinker derived sequence"
FT      54..90
FT      misc_feature
FT      /tag= c
FT      /note= "Residual sequence from the TL-DNA at the
FT      right border repeat"
FT      98..309
FT      3'UTR
FT      /tag= d
FT      /note= "The 3' untranslated end from the TL-DNA
FT      gene 7 (3'g7) of pTiB6S3"
FT      310..331
FT      misc_feature
FT      /tag= e
FT      /note= "Synthetic polylinker derived sequence"
FT      332..883
FT      CDS
FT      /tag= f
FT      /product= "Protein encoded by bialaphos resistance
FT      gene (bar) of Streptomyces hygroscopicus"
FT      884..2609
FT      promoter
FT      /tag= g
FT      /note= "Promoter from the atS1A ribulose-1,5-biphosphate
FT      carboxylase small subunit gene from Arabidopsis thaliana"
FT      2610..2659
FT      misc_feature
FT      /tag= h
FT      /note= "Synthetic polylinker derived sequence"
FT      2660..2920
FT      misc_feature
FT      /tag= i
FT      /note= "TaqI fragment from the 3' untranslated end of the
FT      nopaline synthase gene (3'nos) from the T-DNA of pTiT37
FT      and containing plant polyadenylation signals"
FT      2921..2936
FT      misc_feature
FT      /tag= j
FT      /note= "Synthetic polylinker derived sequence"
FT      2937..3032
FT      3'UTR
FT      /tag= k
FT      /note= "The 3' untranslated region downstream from the
FT      barnase coding sequence of Bacillus amyloliquefaciens"
FT      3033..3368
FT      CDS
FT      /tag= l
FT      /product= "Protein encoded by barnase gene from
FT      Bacillus amyloliquefaciens"
FT      3369..4878
FT      promoter
FT      /tag= m
FT      /note= "Promoter region of the anther-specific gene
FT      TA29 from Nicotiana tabacum"
FT      4879..4924
FT      misc_feature
FT      /tag= n
FT      /note= "Synthetic polylinker derived sequence"
FT      4925..5215
FT      promoter
FT      /tag= o
FT      /note= "Promoter of the nopaline synthase gene from the
FT      T-DNA of pTiT37 of Agrobacterium tumefaciens"
FT      5216..5217
FT      misc_feature
FT      /tag= p
FT      /note= "Synthetic polylinker derived sequence"
FT      5218..5490
FT      CDS
FT      /tag= q
FT      /product= "Protein encoded by barstar gene of
FT      Bacillus amyloliquefaciens"
FT      5491..5530
FT      misc_feature
FT      /tag= r
FT      /note= "Sequence from the 3' untranslated end of the
FT      barstar gene from Bacillus amyloliquefaciens"
FT      5531..5554
FT      misc_feature
FT      /tag= s
FT      /note= "Synthetic polylinker derived sequence"
FT      5555..5766
FT      3'UTR
FT      /tag= t
FT      /note= "The 3' untranslated end from the TL-DNA

```

```

FT      gene 7 (3'g7) of pTiB6S3"
FT      5767..5773
FT      misc_feature
FT      /tag= u
FT      /note= "Synthetic polylinker derived sequence"
FT      5774..5810
FT      misc_feature
FT      /tag= v
FT      /note= "Residual sequence from the TL-DNA at the
FT      right border repeat"
FT      5811..5840
FT      misc_feature
FT      /tag= w
FT      /note= "Synthetic polylinker derived sequence"
FT      5841..5865
FT      misc_feature
FT      /tag= x
FT      /note= "Left border repeat from the TL-DNA from
FT      pTiB6S3"
FT      WO200131042-A2.
FT      XX
FT      PD
FT      03-MAY-2001.
FT      XX
FT      PF
FT      26-OCT-2000; 2000WO-EP10680.
FT      XX
FT      PR
FT      29-OCT-1999; 99US-0430497.
FT      XX
FT      (AVET ) AVENTIS CROPS SCIENCE NV.
FT      PA
FT      Weston B, De Beuckeleer M;
FT      PI
FT      WPI; 2001-300517/31.
FT      XX
FT      Transgenic Brassica plants, seeds, cells or tissues, characterized by
FT      harboring specific transformation events, particularly by presence of
FT      male-sterility gene, at specific location in its genome -
FT      Claim 1; Page 47-49; 53pp; English.
FT      CC
FT      The present invention relates to a transgenic Brassica plant or its
FT      seed, cells or tissues, characterised by harbouring a specific
FT      transformation event, particularly by the presence of a male-sterility
FT      gene, at a specific location in the Brassica genome. Transgenic
FT      Brassica plant is useful for producing a hybrid seed by crossing the
FT      transgenic plant with a male-fertile Brassica plant and harvesting the
FT      hybrid seed from the transgenic Brassica plant.
FT      CC
FT      The present sequence is chimeric T-DNA of plasmid pTCO113. This sequence
FT      comprises right border repeat, left border repeat and 3' untranslated
FT      region (UTR) from TL-DNA of pTiB6S3, synthetic polylinker sequences,
FT      coding regions of bialaphos resistance gene (bar) from
FT      Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens
FT      and barstar gene from Bacillus amyloliquefaciens and promoters of atS1A
FT      ribulose-1,5-biphosphate carboxylase small subunit gene from
FT      Arabidopsis thaliana, the anther-specific gene TA29 from
FT      Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTiT37
FT      of Agrobacterium tumefaciens.
FT      XX
FT      Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;
FT      SQ

```

```

Query Match      46.1%; Score 191.4; DB 22; Length 5865;
Best Local Similarity 99.5%; Pred. No. 1.4e-30;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 223 gatgtacatggccgataaagaaaggcaattttagatgttaattcccatcttgaagaaa 282
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5813 GATGTACATGGTCGATAGCAAAAAGGCAATTTGTAGTGTATTTCCTCCATCTTGAAGAAA 5754
QY 283 tatagtttaataattattgataaataacaagtcaggtagtattatagtcacaagaaaaa 342
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5753 TATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAA 5694
QY 343 taaattattgatgcgaagttaaattcagaataatttcacaaactgattatcagctgg 402
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5693 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGG 5634

```

QY 403 tacattgccgtag 415  
|||||  
Db 5633 TACATTGCCGTAG 5621

## RESULT 5

AAF25320/c  
ID AAF25320 standard; DNA; 7599 BP.

XX  
AC AAF25320;

XX  
DT 30-APR-2001 (first entry)

XX  
DE Nucleotide sequence of a plasmid PGKB5.

XX  
KW Plant promoter; root cell; root-specific expression; parasite resistance;  
KW nematode resistance; fungal resistance; water stress; salt stress;  
KW sugar content; nitrogen transport; ss.

XX  
OS Synthetic.

XX  
PN WO200100833-A1.

XX  
PD 04-JAN-2001.

XX  
PF 23-JUN-2000; 2000WO-FR01768.

XX  
PR 25-JUN-1999; 99FR-0008185.

XX  
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.

XX  
PI Hoffmann B, Mollier P, Pelletier G;

XX  
DR WPI; 2001-102893/11.

XX  
PT New constitutive plant promoter active specifically in roots, useful  
PT for controlling expression of pest or drought resistance genes, and  
PT related transgenic plants -

XX  
PS Disclosure; Fig 9; 92pp; French.

XX  
CC The present sequence represents a plasmid PGKB5. The plasmid contains  
CC a plant promoter that directs expression of a selected sequence in  
CC root cells at all stages of development of a plant. The plant promoter  
CC is used to control expression of genes in a root-specific manner,  
CC especially genes that provide resistance to parasites, pests (nematodes  
CC or fungi), water and salt stress, or alter sugar content or nitrogen  
CC transport. Fragments of the promoter are useful as probes or primers  
CC to detect or amplify at least part of the promoter.

XX  
SQ Sequence 7599 BP; 1972 A; 1938 C; 1937 G; 1752 T; 0 other;

Query Match 46.1%; Score 191.4; DB 22; Length 7599;  
Best Local Similarity 99.5%; Pred. No. 1.4e-30;  
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 223 gatgtacatgccgataagaaagcaattttagatgttaattcccatcttgaagaaa 282  
|||||

Db 6771 GATGTACATGGTCGATTAAGAAAGGCCAATTGTAGATGTTAAATCCCATCTTGAAGAAA 6712  
|||||

QY 283 tatagttaaattattattgataaaataacaagtcagggtattatagtcgaagcaaaaaca 342  
|||||

Db 6711 TATAGTTAAATATTATTATGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAACA 6652  
|||||

QY 343 taaatttattgcaagtttaattcagaataatttcaataactgattatatcagctgg 402  
|||||

Db 6651 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGG 6592  
|||||

QY 403 tacattgccgtag 415  
|||||

Db 6591 TACATTGCCGTAG 6579

## RESULT 6

AAF86439/c  
ID AAF86439 standard; DNA; 5228 BP.

XX  
AC AAF86439;

XX  
DT 25-JUN-2001 (first entry)

XX  
DE Plasmid pTSl72delta.

XX  
KW Male sterile plant; RNAase inhibitor; plasmid pTSl72delta; ds.

XX  
OS Unidentified.

XX  
PN WO200124616-A1.

XX  
PD 12-APR-2001.

XX  
PF 12-SEP-2000; 2000WO-JP06222.

XX  
PR 30-SEP-1999; 99JP-0279307.

XX  
PA (NISB ) JAPAN TOBACCO INC.

XX  
PI Hamada K, Nakakido F;

XX  
DR WPI; 2001-266212/27.

XX  
PT Method for producing male sterile rice and maize by inserting RNAse  
PT gene and RNAse inhibitor genes with promoters into the plant genome -  
XX Disclosure; Page 14-17; 29pp; Japanese.

XX  
CC The present invention relates to a method for producing male sterile  
CC plants. The method comprises inserting a promoter fragment upstream of an  
CC RNAse gene and a second promoter, upstream of an RNAse inhibitor protein  
CC gene and inserting it into the plant genome. The method is useful for  
CC producing male sterile tobacco, lettuce and rapeseed plants, but  
CC preferably rice and maize. The present sequence is a vector used in  
CC the method of the present invention.

XX  
SQ Sequence 5228 BP; 1384 A; 1307 C; 1263 G; 1274 T; 0 other;

Query Match 45.5%; Score 188.8; DB 22; Length 5228;  
Best Local Similarity 99.0%; Pred. No. 4.6e-30;  
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatgcccgaataagaaagcaattttagatgttaattcccatcttgaagaaaat 283  
|||||

Db 5222 ACGTACATGGTCGATAGAAAGGCCAATTGTAGATGTTAATCCCATCTTGAAGAAAT 5163  
|||||

QY 284 atagtttaaatattattgataaaataacaagtcagggtattatagtcgaagcaaaaacat 343  
|||||

Db 5162 ATAGTTTAAATATTATTATTATATAAAATAACAAGTCAGGTATTATTATAGTCCAAAGCAAAACAT 5103  
|||||

QY 344 aaatttttagtcaagtttaattcaaatatttcaataactgattatatcagctgg 403  
|||||

Db 5102 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGGT 5043  
|||||

QY 404 acattgccgtag 415  
|||||

Db 5042 ACATTGCCGTAG 5031

## RESULT 7

AAZ91097/c  
ID AAZ91097 standard; DNA; 6539 BP.

XX  
AC AAZ91097;

XX  
DT 06-JUN-2000 (first entry)





```

PF 03-AUG-1999; 99WO-JP04167.
PR 04-AUG-1998; 98JP-0220060.
XX (NISB ) JAPAN TOBACCO INC.
XX Hamada K, Nakakido F;
XX WPI; 2000-195581/17.
XX Mutate barnase gene for efficient construction of plant transformants,
XX particularly male sterile plants free from any undesirable characters
XX by specifically expressing the gene alone in another -
PS Example 3; Page 19-23; 30pp; Japanese.
XX The invention relates to the generation of male sterile plants by
XX the introduction of a mutant barnase gene (AAZ91095) for expression
XX specifically in the anther of a plant. This sequence represents the
XX E. coli/Agrobacterium shuttle vector plasmid pPS172 which contains
XX the synthetic barnase gene (AAZ91094) under control of the cauliflower
XX mosaic virus 35S promoter. The vector also contains a region of the
XX Agrobacterium T-DNA gene 7. The vector is used for transmitting the
XX barnase gene to plants via an Agrobacterium tumefaciens host cell.
XX The transformed plant is used in plant breeding.
XX Sequence 6548 BP; 1756 A; 1579 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 21; Length 6548;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatgcccgcataagaaggaagcaattttagatgttattcccatcttgaagaagaat 283
Db 6542 ACGTACATGGTTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTGAAGAAAT
QY 284 atagtttaaatatttattgataaaatacaagtcaggtattattcgaactgattatcagctggt 343
Db 6482 ATAGTTTAAATATTATTATGATAAAATAACAAGTCAGGTATTATAGTCCAGCAAAACAT
QY 344 aaatttattgacgaagtttaaatcagaataatttcgaactgattatcagctggt 403
Db 6422 AAATTTATTGTCGAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 6363
QY 404 acattgcccgtag 415
Db 6362 ACATTGCCGTAG 6351

RESULT 11
AAF86441/C
ID AAF86441 standard; DNA; 7492 BP.
XX AAF86441;
AC AAF86441;
XX 25-JUN-2001 (first entry)
XX Plasmid pTS346.
XX Male sterile plant; RNAase inhibitor; plasmid pTS346; ds.
XX Unidentified.
XX WO200124616-A1.
XX 12-APR-2001.
XX 12-SEP-2000; 2000WO-JP06222.
XX 30-SEP-1999; 99JP-0279307.
XX (NISB ) JAPAN TOBACCO INC.

```

```

XX Hamada K, Nakakido F;
XX WPI; 2001-266212/27.
XX Method for producing male sterile rice and maize by inserting RNAse
XX gene and RNAse inhibitor genes with promoters into the plant genome -
XX Disclosure; Page 19-23; 29pp; Japanese.
XX The present invention relates to a method for producing male sterile
XX plants. The method comprises inserting a promoter fragment upstream of an
XX RNAse gene and a second promoter, upstream of an RNAse inhibitor protein
XX gene and inserting it into the plant genome. The method is useful for
XX producing male sterile tobacco, lettuce and rapeseed plants, but
XX preferably rice and maize. The present sequence is a vector used in
XX the method of the present invention.
XX Sequence 7492 BP; 1987 A; 1801 C; 1752 G; 1952 T; 0 other;

Query Match 45.5%; Score 188.8; DB 22; Length 7492;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatgcccgcataagaaggaagcaattttagatgttattcccatcttgaagaagaat 283
Db 7486 ACGTACATGGTTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTGAAGAAAT 7427
QY 284 atagtttaaatatttattgataaaatacaagtcaggtattattcgaactgattatcagctggt 343
Db 7426 ATAGTTTAAATATTATTATGATAAAATAACAAGTCAGGTATTATAGTCCAGCAAAACAT 7367
QY 344 aaatttattgacgaagtttaaatcagaataatttcgaactgattatcagctggt 403
Db 7366 AAATTTATTGTCGAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 7307
QY 404 acattgcccgtag 415
Db 7306 ACATTGCCGTAG 7295

RESULT 12
AAT39337/C
ID AAT39337 standard; DNA; 1303 BP.
XX AAT39337;
AC AAT39337;
XX 22-JAN-1997 (first entry)
XX Plasmid pTS88 (EcoRI-HindIII fragment).
XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
XX transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX Synthetic.
XX Key Location/Qualifiers
XX misc_feature 1..35
FT /tag= a
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
FT /tag= b
FT /label= p35S
FT /function= 35S promoter of cauliflower mosaic virus
FT 695..967
FT /tag= c
FT /label= barstar
FT /product= Bacillus amyloliquefaciens barstar
FT 968..1287
FT polyA_signal
FT /tag= d

```



```
FT /label= 3'g7
FT /function= region containing polyadenylation signal
FT of gene 7 og Agrobacterium T-DNA
FT misc_feature 1288..1303
FT /*tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX
XX WO9626283-A1.
XX
XX 29-AUG-1996.
XX
XX 21-FEB-1996; 96WO-EP00722.
XX
XX 21-FEB-1995; 95EP-0400364.
XX
XX (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
XX Botterman J, Cornelissen M, Michiels F;
XX
XX WPI; 1996-402373/40.
XX
XX Prodn. of male sterile plants by transforming with a chimaeric
XX construct - comprising a male sterility DNA e.g. barnase and a
XX co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX for generating hybrid cultivars
XX
XX Example 1; Page 38; 56pp; English.
XX
XX The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains
XX barstar DNA under control of a 35S promoter. The plasmid was
XX used with pTS174 (see also AAT39336) contg. barnase DNA under
XX control of the stamen-specific promoter E1 to produce male sterile
XX rice cv. Kochihibiki transgenic plants, and with plasmid pVE136
XX (see also AAT39338) contg. barnase DNA under control of the stamen-
XX specific PCA55 promoter to produce male sterile maize plants.
XX Expression of barnase (a ribonuclease) in the stamen leads to male
XX sterility. Constitutive expression of barstar counteracts possible
XX low level expression of barnase DNA in non-stamen tissue.
XX
XX Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 45.4%; Score 188.4; DB 17; Length 1303;
Best Local Similarity 99.5%; Pred. No. 5e-30;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatggccgataaagaaagcaatttggtagattgtaattcccatcttgaagaataat 285
Db 1287 GTACATGTCGATGAAGAAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAATAT 1228

Qy 286 agtttaaatatttattgataaataacaagtcaggtattattagtcacgaacaaacataa 345
Db 1227 AGTTTAAATATTATTGATAAATAACAAGTCAGGTATTATATGCCAAGCAAAACATATA 1168

Qy 346 atttattgatgcaagtttaattcagaataatttcaataactgattatcagctggtag 405
Db 1167 ATTTATTGATCAGTTTAATTTCAGAAATATTTCATTAACATGATTATATCAGCTGGTAC 1108

Qy 406 attgccgtag 415
Db 1107 ATTGCCGTAG 1098

RESULT 13
AAQ14529/c
ID AAQ14529 standard; DNA; 3201 BP.
XX
XX AC AAQ14529;
XX
XX 27-JAN-1992 (first entry)
XX
XX pPS029 Bt ICP coding sequence.
```

```
XX Bacillus thuringiensis; insecticidal crystal protein; ICP;
KW deletion; ss.
XX Synthetic.
XX WO9116432-A.
XX
XX 31-OCT-1991.
XX
XX 17-APR-1991; 91WO-EP00733.
XX
XX 18-APR-1990; 90EP-0401055.
XX
XX (PLAN-) PLANT GENETIC SYST.
XX
XX Cornelissen M, Soetaert P, Stam M, Dockx J;
XX
XX WPI; 1991-339820/46.
XX
XX Modified Bacillus thuringiensis insecticidal crystal protein
XX genes - having A and T sequences changed to G and C sequences
XX encoding same amino acids, for increased expression levels
XX
XX Disclosure; Fig 6(c); 78pp; English.
XX
XX "n" in the sequence refers to not known nucleotides.
XX
XX pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-
XX terminal modification and the internal modification of the Bt ICP
XX coding sequence.
XX See also AAQ14529, AAQ15142-44.
XX
XX Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;

Query Match 45.4%; Score 188.4; DB 12; Length 3201;
Best Local Similarity 99.5%; Pred. No. 5.3e-30;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatggccgataaagaaagcaatttggtagattgtaattcccatcttgaagaataat 285
Db 3141 GTACATGTCGATGAAGAAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAATAT 3082

Qy 286 agtttaaatatttattgataaataacaagtcaggtattattagtcacgaacaaacataa 345
Db 3081 AGTTTAAATATTATTGATAAATAACAAGTCAGGTATTATATGCCAAGCAAAACATATA 3022

Qy 346 atttattgatgcaagtttaattcagaataatttcaataactgattatcagctggtag 405
Db 3021 ATTTATTGATCAGTTTAATTTCAGAAATATTTCATTAACATGATTATATCAGCTGGTAC 2962

Qy 406 attgccgtag 415
Db 2961 ATTGCCGTAG 2952

RESULT 14
AAH25423
ID AAH25423 standard; DNA; 4832 BP.
XX
XX AC AAH25423;
XX
XX 22-AUG-2001 (first entry)
XX
XX Nucleotide sequence of plasmid pTHW118.
XX
XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barstar gene; ss.
XX
XX Synthetic.
XX Streptomyces hygroscopicus.
XX Arabidopsis thaliana.
XX Bacillus amyloliquefaciens.
```



```
FT CDS complement (331..882)
FT /tag= c
FT /label= Bar
FT /product= phosphinothricin acetyltransferase
FT promoter complement (883..2608)
FT /tag= d
FT /label= PSSU
FT /note= "promoter region of Rubisco small subunit
FT gene of Arabidopsis thaliana"
FT 3'UTR complement (2658..3031)
FT /tag= e
FT /label= 3'nos
FT /note= "3'untranslated region contg. the poly-A
FT signal of the nopaline-synthase gene of
FT Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /tag= f
FT /label= Barnase
FT /product= barnase
FT promoter complement (3368..4876)
FT /tag= g
FT /label= PTA29
FT /note= "promoter region of tobacco TA29 gene"
FT misc_RNA complement (4822..4946)
FT /tag= h
FT /label= LB
FT /note= "T-DNA left border"
XX
XX EP757102-A1. PN
XX
XX 05-FEB-1997. XX
XX
XX 04-AUG-1995; 95EP-0401844. XX
XX
XX 04-AUG-1995; 95EP-0401844. XX
XX (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
XX De Block M;
XX
XX WPI; 1997-111050/11. XX
XX
XX Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase
XX inhibitor - reduces the cultured cells response to stress and
XX reduces metabolism
XX
XX Example 3; Page 13-16; 25pp; English.
XX
XX Plasmid pTHW107 is a vector carrying T-DNA (AAT59531) comprising a
XX barnase coding sequence under control of the tobacco TA29 gene
XX stamen-specific promoter and a phosphinothricin acetyltransferase
XX coding sequence under control of an Arabidopsis Rubisco small
XX subunit gene promoter. Oilseed rape hypocotyl explants were
XX infected with Agrobacterium tumefaciens C58ClRif carrying vector
XX pTHW107 and helper Ti plasmid pMP60. In some treatments, the
XX hypocotyls were treated with the poly-(ADP-ribose) polymerase
XX inhibitor niacinamide (250 mg/l) 4 days prior to infection.
XX Plants regenerated from niacinamide-treated transformed calli
XX had a low copy number and displayed less variation in the
XX expression profile of the transgenes.
XX
XX Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;
```

```
Query Match 45.4%; Score 188.4; DB 18; Length 4946;
Best Local Similarity 99.5%; Pred. No. 5.5e-30;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 226 gtacatgccatgaagaaaggcattgttagattgtaattcccatcttgaaagaaatat 285
Db 54 gtacatggctcgaagaagaaaggcattgttagattgtaattcccatcttgaaagaaatat 113
Oy 286 agtttaaatattattgataaataacaagtcaggtattattagtcctcaagcaaaacataa 345
```

Search completed: February 15, 2002, 19:01:52  
Job time: 20756 sec

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:07:35 ; Search time 353.79 Seconds  
(without alignments)  
265.661 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagtttggtgtcatga.....cagctgtgacattgcgtag 415

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PT05\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	191.4	46.1	5864	3	US-08-894-440-4 Sequence 4, Appli
C 2	188.8	45.5	6548	3	US-08-894-440-1 Sequence 1, Appli
C 3	188.8	45.5	6548	3	US-08-817-188-2 Sequence 2, Appli
C 4	188.8	45.5	7811	2	US-08-549-680A-5 Sequence 5, Appli
C 5	188.4	45.4	1303	3	US-08-894-440-2 Sequence 2, Appli
C 6	188.4	45.4	3200	1	US-08-453-104-23 Sequence 23, Appli
C 7	188.4	45.4	3200	2	US-08-694-824-23 Sequence 23, Appli
C 8	188.4	45.4	4946	3	US-08-817-188-1 Sequence 1, Appli
C 9	188.4	45.4	5560	3	US-08-817-188-5 Sequence 5, Appli
C 10	188.4	45.4	7566	2	US-08-232-016-23 Sequence 23, Appli
C 11	188.4	45.4	7639	3	US-08-232-016-22 Sequence 22, Appli
C 12	186.8	45.0	5864	3	US-08-894-440-4 Sequence 4, Appli
C 13	177	42.7	3201	1	US-08-453-104-22 Sequence 22, Appli
C 14	177	42.7	3201	2	US-08-694-824-22 Sequence 22, Appli
C 15	153	36.9	1186	1	US-08-064-121-2 Sequence 2, Appli
C 16	153	36.9	1186	3	US-08-478-015-2 Sequence 2, Appli
C 17	153	36.9	1186	3	US-08-475-975-2 Sequence 2, Appli
C 18	153	36.9	1186	3	US-09-084-889-2 Sequence 2, Appli
C 19	146	35.2	3153	4	US-09-080-625-3 Sequence 3, Appli
C 20	146	35.2	3336	4	US-09-080-625-2 Sequence 2, Appli
C 21	146	35.2	3694	4	US-09-080-625-5 Sequence 5, Appli
C 22	146	35.2	3877	4	US-09-080-625-4 Sequence 4, Appli
C 23	143	34.5	24595	6	5428147-1 Patent No. 5428147
C 24	44.8	10.8	8654	1	US-08-920-812-6 Sequence 6, Appli
C 25	44.8	10.8	8654	1	US-08-920-812-6 Sequence 6, Appli
C 26	44.8	10.8	8654	1	US-08-921-177-6 Sequence 6, Appli
C 27	44.8	10.8	8654	1	US-08-362-577C-6 Sequence 6, Appli

28	44.8	10.8	8654	2	US-08-920-828-6	Sequence 6, Appli
C 29	44.6	10.7	5526	3	US-08-751-359-21	Sequence 21, Appli
C 30	43.6	10.5	19124	2	US-08-487-826B-13	Sequence 13, Appli
C 31	43	10.4	19124	2	US-08-487-826B-13	Sequence 13, Appli
C 32	42.6	10.3	8920	2	US-08-446-855A-1	Sequence 1, Appli
C 33	42.6	10.3	8920	4	US-09-150-741-1	Sequence 1, Appli
C 34	42	10.1	1316	2	US-08-871-924A-1	Sequence 1, Appli
C 35	40.4	9.7	1588	3	US-09-058-489-45	Sequence 45, Appli
C 36	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appli
C 37	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appli
C 38	39.8	9.6	24595	6	5428147-1	Patent No. 5428147
C 39	39.4	9.5	80595	4	US-09-078-294-3	Sequence 3, Appli
C 40	39.2	9.4	658	4	US-08-998-416-595	Sequence 595, App
C 41	38.8	9.3	636	4	US-08-998-416-1137	Sequence 1137, Ap
C 42	38.8	9.3	660	1	US-07-991-867B-32	Sequence 32, Appli
C 43	38.8	9.3	660	1	US-08-107-755A-32	Sequence 32, Appli
C 44	38.8	9.3	660	2	US-08-544-332-32	Sequence 32, Appli
C 45	38.8	9.3	1511	1	US-07-991-867B-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-08-894-440-4/c  
; Sequence 4, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894,440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 5864  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; OTHER INFORMATION: plasmid pCOL13  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((1)..(25))  
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((98)..(330))  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((331)..(882))  
; OTHER INFORMATION: region coding for phosphinothricin acetyl  
; OTHER INFORMATION: transferase (bar)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((883)..(2608))  
; OTHER INFORMATION: promoter of small  
; OTHER INFORMATION: subunit gene of Rubisco of  
; OTHER INFORMATION: Arabidopsis (Pssu)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2659)..(3031))  
; OTHER INFORMATION: region containing  
; OTHER INFORMATION: polyadenylation signal of  
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA  
; OTHER INFORMATION: (3'nos)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((3032)..(3367))  
; OTHER INFORMATION: region coding for  
; OTHER INFORMATION: barnase of Bacillus  
; OTHER INFORMATION: amyloliquefaciens  
; FEATURE:



; LOCATION: Complement((2019)..(2288))  
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
; OTHER INFORMATION: gene of Agrobacterium T-DNA.  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2289)..(2624))  
; OTHER INFORMATION: barnase: region coding for barnase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2625)..(4313))  
; OTHER INFORMATION: P1: promoter region of E1 gene of rice  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4336)..(5170)  
; OTHER INFORMATION: P35S: 35S promoter region of Cauliflower Mosaic  
; OTHER INFORMATION: Virus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((5711)..(6262))  
; OTHER INFORMATION: bar: region coding for phosphinotricin  
; OTHER INFORMATION: acetyltransferase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (6263)..(6496)  
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
; OTHER INFORMATION: T-DNA  
US-08-817-188-2

Query Match 45.5%; Score 188.8; DB 3; Length 6548;  
Best Local Similarity 99.0%; Pred. No. 1.1e-36;  
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 224 atgtacatggccgataaagaaggaattttagatgttaattcccatcttgaagaagaat 283  
Db 6542 ACGTACATGGTCGATAGAAAAGGCAATTTGTAGATGTTAATTCCTCCATCTTGAAGAAGAAAT 6483  
  
Qy 284 atagttaataattattgataaaaatacaagtcaggatgattatagtcacaaagcaaaaacat 343  
Db 6482 ATAGTTTAATATATTTATTGATAAAATAACAGTCAGGTATTATATATCCAAAGCAAAACAT 6423  
  
Qy 344 aaattattgatgcaggttaaaattcagaataatttcataaactgattatcagctggt 403  
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATRACTGATTATATATCAGCTGCT 6363  
  
Qy 404 acattgccgtag 415  
Db 6362 ACATTGCCGTAG 6351

RESULT 4  
US-08-549-680A-5  
; Sequence 5, Application US/08549680A  
; Patent No. 5962768  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISEN, MARCUS  
; APPLICANT: REYNAERTS, ARLETTE  
; APPLICANT: GOSSELE, VERONIQUE  
; APPLICANT: VAN AARSSEN, ROEL  
; TITLE OF INVENTION: MARKER GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/549,680A  
FILING DATE: 16 JANUARY 1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-0111P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7811 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (synthetic)  
FEATURE:  
NAME/KEY: misc\_recomb  
LOCATION: 1..7811  
OTHER INFORMATION: /label= vector ptrVA3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 194..218  
OTHER INFORMATION: /note= "T-DNA right border"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 484..684  
OTHER INFORMATION: /note= "the 3' end formation and  
OTHER INFORMATION: polyadenylation region of T-DNA gene 7"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (729..1340)  
OTHER INFORMATION: /note= "the aac(6') coding  
OTHER INFORMATION: sequence"  
FEATURE:  
NAME/KEY: promoter  
LOCATION: 1341..1756  
OTHER INFORMATION: /label= 35S promoter  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3001..3023  
OTHER INFORMATION: /note= "T-DNA left border  
OTHER INFORMATION: sequences"  
US-08-549-680A-5

Query Match 45.5%; Score 188.8; DB 2; Length 7811;  
Best Local Similarity 99.0%; Pred. No. 1.2e-36;  
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 224 atgtacatggccgataaagaaggaattttagatgttaattcccatcttgaagaagaat 283  
Db 438 ACGTACATGGTCGATAGAAAAGGCAATTTGTAGATGTTAATTCCTCCATCTTGAAGAAGAAAT 497  
  
Qy 284 atagttaataattattgataaaaatacaagtcaggatgattatagtcacaaagcaaaaacat 343  
Db 498 ATAGTTTAAATATATTTATTGATAAAATAACAGTCAGGTATTATATAGTCCAAAGCAAAACAT 557  
  
Qy 344 aaattattgatgcaggttaaaattcagaataatttcataaactgattatcagctggt 403  
Db 558 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATRACTGATTATATATCAGCTGCT 617  
  
Qy 404 acattgccgtag 415  
Db 618 ACATTGCCGTAG 629  
  
RESULT 5

US-08-894-440-2/c

; Sequence 2, Application US/08894440

; Patent No. 6025546

; GENERAL INFORMATION:

; APPLICANT: PLANT GENETIC SYSTEMS N.V.

; TITLE OF INVENTION: Method to obtain male sterile plants

; FILE REFERENCE: NMSCOR

; CURRENT APPLICATION NUMBER: US/08/894,440

; CURRENT FILING DATE: 1997-11-12

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 1303

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI

; OTHER INFORMATION: fragment of pTS88

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)..(35)

; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (36)..(694)

; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain

; OTHER INFORMATION: CM1841 (P35S)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (695)..(967)

; OTHER INFORMATION: region coding for barstar of Bacillus

; OTHER INFORMATION: amyloliquefaciens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (968)..(1287)

; OTHER INFORMATION: region containing polyadenylation signal of gene 7

; OTHER INFORMATION: of Agrobacterium T-DNA (3'5')

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1288)..(1303)

; OTHER INFORMATION: polylinker of pGEM2

US-08-894-440-2

Query Match

Best Local Similarity 45.4%; Score 188.4; DB 3; Length 1303;

Mismatches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatggccgataaagaaggcaattttagatgttaattcccatcttgaagaataat 285

Db 1287 GTACATGGTTCGATAAGAAAGGCAATTGTAGATGTTAATTCCTCTTGAAGAAATAT 1228

QY 286 agtttaaatattattatgataaataacaagtcaggtagattatagtcacaaacataa 345

Db 1227 AGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCAAGCAAAACATAA 1168

QY 346 atttattgatgcaagtttaattcagaataatttcaataactgattatatacagctggtag 405

Db 1167 ATTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATCAGCTGGTAC 1108

QY 406 attgocgtag 415

Db 1107 ATTGCGGTAG 1098

RESULT 6

US-08-453-104-23/c

; Sequence 23, Application US/08453104

; Patent No. 5633446

; GENERAL INFORMATION:

; APPLICANT: CORNELIJSSEN, Marc

; APPLICANT: SOETABERT, Piet

; APPLICANT: STAM, Maïke

; APPLICANT: DOCKX, Jan

; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS

; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: George Mason Bldg., Washington & Prince Sts.

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/453,104

; FILING DATE:

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/937,869

; FILING DATE: 16-DEC-1992

; APPLICATION NUMBER: GB 90401055.0

; FILING DATE: 18-APR-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Rea, Teresa S

; REGISTRATION NUMBER: 30,427

; REFERENCE/DOCKET NUMBER: 010830-032

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3200 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 2078..2082

; OTHER INFORMATION: /note= "Nucleotides 2078-2082

; OTHER INFORMATION: wherein N is not known."

US-08-453-104-23

Query Match

Best Local Similarity 45.4%; Score 188.4; DB 1; Length 3200;

Mismatches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatggccgataaagaaggcaattttagatgttaattcccatcttgaagaataat 285

Db 3141 GTACATGGTTCGATAAGAAAGGCAATTGTAGATGTTAATTCCTCTTGAAGAAATAT 3082

QY 286 agtttaaatattattatgataaataacaagtcaggtagattatagtcacaaacataa 345

Db 3081 AGTTTAAATATTATTGATAAATAACAAGTCAGGTATTATAGTCCAAGCAAAACATAA 3022

QY 346 atttattgatgcaagttttaaattcagaataatttcaataactgattatatacagctggtag 405

Db 3021 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGTAC 2962

QY 406 attgocgtag 415

Db 2961 ATTGCGGTAG 2952

RESULT 7

US-08-694-824-23/c

; Sequence 23, Application US/08694824

; Patent No. 5877306



GENERAL INFORMATION:  
APPLICANT: CORNELISSEN, Marc  
APPLICANT: SOETAERT, Piet  
APPLICANT: STAM, Janke  
APPLICANT: DOCKX, Jan  
TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
TITLE OF INVENTION: IN PLANT CELLS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/694,824  
FILING DATE: 09-AUG-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,869  
FILING DATE: 16-DEC-1992  
APPLICATION NUMBER: GB 90401055.0  
FILING DATE: 18-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Rea, Teresa S  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 010830-032  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3200 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2078..2082  
OTHER INFORMATION: /note= "Nucleotides 2078-2082  
OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 45.4%; Score 188.4; DB 2; Length 3200;  
Best Local Similarity 99.5%; Pred. No. 1.3e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatggccgataaagaaagcaattgttagatgtttaattcccatcttgaagaataat 285  
Db 3141 GTACATGGTCGATGAAGAAAGGCAATTGTAGATGTTAATTCCTTCCCAAGAAATAT 3082  
QY 286 agtttaaatatttattgataaaatacaagtcagggtattatagtcacagcaaaaacataa 345  
Db 3081 AGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAACATAA 3022  
QY 346 atttatgtatgaagtttaattcagaataatttcataataactgatttatcagctgttac 405  
Db 3021 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGTAC 2962  
QY 406 attgccgtag 415  
Db 2961 ATTGCCGTAG 2952

RESULT 8  
US-08-817-188-1  
Sequence 1, Application US/08817188  
Patent No. 6074876  
GENERAL INFORMATION:  
APPLICANT: DE BLOCK, MARC  
TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
FILE REFERENCE: 2121-0127P  
CURRENT APPLICATION NUMBER: US/08/817,188  
CURRENT FILING DATE: 1997-05-15  
EARLIER APPLICATION NUMBER: PCT/EP96/03366  
EARLIER FILING DATE: 1996-07-31  
EARLIER APPLICATION NUMBER: EP 95401844.6  
EARLIER FILING DATE: 1995-08-04  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 4946  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
OTHER INFORMATION: plasmid pTHW107  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((1)..(25))  
OTHER INFORMATION: T-DNA right border (RB)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((97)..(330))  
OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
OTHER INFORMATION: T-DNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((331)..(882))  
OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
OTHER INFORMATION: transferase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((883)..(2608))  
OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((2658)..(3031))  
OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
OTHER INFORMATION: gene of Agrobacterium T-DNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((3032)..(3367))  
OTHER INFORMATION: barnase: region coding for barnase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((3368)..(4876))  
OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana  
OTHER INFORMATION: tabacum  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((4922)..(4946))  
OTHER INFORMATION: LB: T-DNA left border  
US-08-817-188-1

Query Match 45.4%; Score 188.4; DB 3; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 1.1e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatggccgataaagaaagcaattgttagatgtttaattcccatcttgaagaataat 285  
Db 54 gtacatggctcgaagaaagcaattgttagatgtttaattcccatcttgaagaataat 113

```
QY 286 agtttaaatatttattgataataaacaagtcagggtattattagtcacaagcaaaaacataa 345
Db 114 agtttaaatatttattgataataaacaagtcagggtattattagtcacaagcaaaaacataa 173
QY 346 atttattgacgaagtttaattcagaataatttcaataactgattatcagctgggtac 405
Db 174 atttattgacgaagtttaattcagaataatttcaataactgattatcagctgggtac 233
QY 406 attgcccgtag 415
Db 234 attgcccgtag 243

RESULT 9
US-08-817-188-5
; Sequence 5, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTIb6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' q7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: pssu: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-Ls1 gene
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (5067)..(5502)
; OTHER INFORMATION: p35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTIb6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; US-08-817-188-5

Query Match 45.4%; Score 188.4; DB 3; Length 5560;
Best Local Similarity 99.5%; Pred. No. 1.4e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatggccgataaagaaagcaatttttagatgtttaattcccatctctgaagaataat 285
Db 41 gtacatggccgataaagaaagcaatttttagatgtttaattcccatctctgaagaataat 100
QY 286 agtttaaatatttattgataataaacaagtcagggtattattagtcacaagcaaaaacataa 345
Db 101 agtttaaatatttattgataataaacaagtcagggtattattagtcacaagcaaaaacataa 160
QY 346 atttattgacgaagtttaattcagaataatttcaataactgattatcagctgggtac 405
Db 161 atttattgacgaagtttaattcagaataatttcaataactgattatcagctgggtac 220
QY 406 attgcccgtag 415
Db 221 attgcccgtag 230

RESULT 10
US-08-232-016-23/c
; Sequence 23, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; NUMBER OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016
```

Db	2070	GTACATGGTCGATGAAGAAAAGGCAATTGTGAGATGTTAAATTCCTCATCTTTGAAGAAATAT	201
Qy	286	agtttaaatatttattgataaaataacaagtcaggtattatagtcacaaagcaaaaacataa	345
Db	2010	AGTTTAAATATTTATTGATAAATAACAAGTCAGGTATTATAGTCCACGAAAAACATAA	195
Qy	346	atttattgatcgaagtttaattcaaatatttcgaataatttcgaataactgattatatacsgctggtag	405
Db	1950	ATTTTATTGATCGCAAGTTTAAATTCAGAAAATATTTCATAACTGATTATATATCAGCTGGTAC	189
Qy	406	attgccgtag	415
Db	1890	ATTCGGTAG	1881
RESULT 11			
US-08-232-016-22/c			
; Sequence 22, Application US/08232016			
; Patent No. 5952547			
; GENERAL INFORMATION:			
; APPLICANT: CORNELISSEN, Marc			
; APPLICANT: SOETAERT, Piet			
; APPLICANT: STAM, Maïke			
; APPLICANT: DOCKX, Jan			
; APPLICANT: VAN ARSEN, Roel			
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN			
; PLANT OF INVENTION: PLANT CELLS			
; NUMBER OF SEQUENCES: 23			
; CORRESPONDENCE ADDRESSES:			
; ADDRESSEE: Burns, Doane, Swecker & Mathis			
; STREET: P.O. Box 1404			
; CITY: Alexandria			
; STATE: Virginia			
; COUNTRY: United States			
; ZIP: 22313-1404			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/232,016			
; FILING DATE: 03-NOV-1994			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: FR 91402920.2			
; FILING DATE: 30-OCT-1991			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: GB 92400820.4			
; FILING DATE: 25-MAR-1992			
; ATTORNEY/AGENT INFORMATION:			
; NAME: McGowan, Malcolm K			
; REGISTRATION NUMBER: F39,300			
; REFERENCE/DOCKET NUMBER: 010830-049			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (703) 836-6620			
; TELEFAX: (703) 836-2021			
; INFORMATION FOR SEQ ID NO: 22:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 7639 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: circular			
; MOLECULE TYPE: DNA (genomic)			
; ORIGINAL SOURCE:			
; ORGANISM: plasmid DNA designated as pJD884			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 1..1869			
; OTHER INFORMATION: /note= "Coding region of a			
; OTHER INFORMATION: truncated bt2 (cryIab) gene, also designated as			
; FEATURE:			
; NAME/KEY: misc feature			

```

; LOCATION: 1877..2110
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2480..3005
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3006..3665
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3666..4491
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5684..6541
; OTHER INFORMATION: /note= "Sequence complementary to
; Patent No. 5952547
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7155..7639
; OTHER INFORMATION: /note= "TR1' and TR2' promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA."
; US-08-232-016-22

```

```

Query Match 45.4%; Score 188.4; DB 2; Length 7639;
Best Local Similarity 99.5%; Pred. No. 1.5e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgaataagaaaggcaattttagatgttaattccatcttgaagaataat 285
Db 2154 GTACATGTCGATAAGAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAAATAT 2095
QY 286 agtttaaatattattgataaaataacaaagtcaggtattatagtcacaaagcaaaacataa 345
Db 2094 AGTTTAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAACATAA 2035
QY 346 atttattgatcgaagtttaattcagaataatttcataactgattatcagctggtac 405
Db 2034 ATTTATGATGCAAGTTTAATTTCAGAAATATTTCAATAACTGATTATATCATCAGCTGGTAC 1975
QY 406 attgcgctag 415
Db 1974 ATTGCCGTAG 1965

```

```

RESULT 12
US-08-894-440-4
; Sequence 4, Application US/08894440
; Patent No. 6025346
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of

```

```

; OTHER INFORMATION: plasmid pTCO113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
; US-08-894-440-4

```

```

Query Match 45.0%; Score 186.8; DB 3; Length 5864;
Best Local Similarity 98.9%; Pred. No. 3.4e-36;
Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 226 gtacatgcccgaataagaaaggcaattttagatgttaattccatcttgaagaataat 285
Db 54 gtacatgctcgataagaaaggcaattttagatgttaattccatcttgaagaataat 113
QY 286 agtttaaatattattgataaaataacaaagtcaggtattatagtcacaaagcaaaacataa 345
Db 114 agtttaaatattattgataaaataacaaagtcaggtattatagtcacaaagcaaaacataa 173
QY 346 atttattgatcgaagtttaattcagaataatttcataactgattatcagctggtac 405
Db 174 atttattgatcgaagtttaattcagaataatttcataactgattatcagctggtac 233

```

Qy 406 attgccgtag 415  
 Db 234 attgccgtag 243

RESULT 13

US-08-453-104-22/c  
 ; Sequence 22, Application US/08453104  
 ; Patent No. 5633446  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CORNELISSEN, Marc  
 ; APPLICANT: SOETAERT, Piet  
 ; APPLICANT: STAM, Maïke  
 ; APPLICANT: DOCKX, Jan  
 ; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
 ; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
 ; TITLE OF INVENTION: IN PLANT CELLS  
 ; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Burns, Doane, Swecker & Mathis  
 ; STREET: George Mason Bldg., Washington & Prince Sts.  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: United States  
 ; ZIP: 22313-1404  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/453,104  
 ; FILING DATE:  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/937,869  
 ; FILING DATE: 16-DEC-1992  
 ; APPLICATION NUMBER: GB 90401055.0  
 ; FILING DATE: 18-APR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rea, Teresa S  
 ; REGISTRATION NUMBER: 30,427  
 ; REFERENCE/DOCKET NUMBER: 010830-032  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 836-6620  
 ; TELEFAX: (703) 836-2021  
 ; INFORMATION FOR SEQ ID NO: 22:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3201 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 2151..2155  
 ; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
 ; OTHER INFORMATION: wherein N is not known."  
 US-08-453-104-22

Query Match 42.7%; Score 177; DB 1; Length 3201;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
 Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 239 aagaaaggcaatttgtagtgaattcccatcttgaaagaataatagtttaaatattt 298  
 Db 3201 AAGAAAGGCAATTGTAGATTAAATCCCATCTTGAAGAAATATAGTTAAATATT 3142

Qy 299 attgataaaatacaagtcaggattatagtcgaagcaaaaacataattttattgatgca 358  
 Db 3141 ATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACATAAAATTTATTGATGCA 3082

Qy 359 agtttaaaattcagaataatttcaataactgattatcagctggtagacattgccgtag 415  
 Db 3081 AGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGGTACATTGCCGTAG 3025

RESULT 14

US-08-694-824-22/c  
 ; Sequence 22, Application US/08694824  
 ; Patent No. 5877306  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CORNELISSEN, Marc  
 ; APPLICANT: SOETAERT, Piet  
 ; APPLICANT: STAM, Maïke  
 ; APPLICANT: DOCKX, Jan  
 ; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
 ; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
 ; TITLE OF INVENTION: IN PLANT CELLS  
 ; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Burns, Doane, Swecker & Mathis  
 ; STREET: George Mason Bldg., Washington & Prince Sts.  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: United States  
 ; ZIP: 22313-1404  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/694,824  
 ; FILING DATE: 09-AUG-1996  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/937,869  
 ; FILING DATE: 16-DEC-1992  
 ; APPLICATION NUMBER: GB 90401055.0  
 ; FILING DATE: 18-APR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rea, Teresa S  
 ; REGISTRATION NUMBER: 30,427  
 ; REFERENCE/DOCKET NUMBER: 010830-032  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 836-6620  
 ; TELEFAX: (703) 836-2021  
 ; INFORMATION FOR SEQ ID NO: 22:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3201 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 2151..2155  
 ; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
 ; OTHER INFORMATION: wherein N is not known."  
 US-08-694-824-22

Query Match 42.7%; Score 177; DB 2; Length 3201;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
 Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 239 aagaaaggcaatttgtagtgaattcccatcttgaaagaataatagtttaaatattt 298  
 Db 3201 AAGAAAGGCAATTGTAGATTAAATCCCATCTTGAAGAAATATAGTTAAATATT 3142

Qy 299 attgataaaatacaagtcaggattatagtcgaagcaaaaacataattttattgatgca 358  
 Db 3141 ATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACATAAAATTTATTGATGCA 3082



---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:03:55 ; Search time 9904.61 Seconds  
(without alignments)  
450.245 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagtttggttgcata.....cagctgtacattgcgctag 415

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estfun:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estom:\*
  - 5: em\_estpl:\*
  - 6: em\_estba:\*
  - 7: em\_estro:\*
  - 8: em\_estov:\*
  - 9: em\_htc:\*
  - 10: gb\_est1:\*
  - 11: gb\_est2:\*
  - 12: gb\_hc:\*
  - 13: gb\_gss:\*
  - 14: em\_gss\_fun:\*
  - 15: em\_gss\_hum:\*
  - 16: em\_gss\_inv:\*
  - 17: em\_gss\_pln:\*
  - 18: em\_gss\_pro:\*
  - 19: em\_gss\_rod:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	63	15.2	734	13	CNS010MP	AL099163 Drosophill
c 2	61	14.7	1101	13	CNS0042W	AL055440 Drosophill
c 3	57.2	13.8	844	13	CNS03D0I	AL238491 Tetraodon
c 4	57.2	13.8	1101	13	CNS0008B	AL063632 Drosophill
c 5	57.2	13.8	1101	13	CNS00238	AL097166 Drosophill
c 6	55.4	13.3	1101	13	CNS016LI	AL106896 Drosophill
c 7	55	13.3	537	13	AQ506817	AQ506817 RPCI-11-2
c 8	54.8	13.2	987	13	CNS014PQ	AL104456 Drosophill
c 9	54.8	13.2	1101	13	CNS003BD	AL064091 Drosophill
c 10	54.6	13.2	980	13	CNS00JG1	AL076232 Drosophill
c 11	54.6	13.2	1101	13	CNS00EVL	AL069706 Drosophill
c 12	54.4	13.1	893	13	CNS013XE	AL103436 Drosophill

13	54	13.0	905	13	CNS00KHX	AL077798 Drosophill
c 14	53.4	12.9	500	10	AU087444	AU087444 Drosophill
c 15	53.4	12.9	970	13	CNS0182A	AL108796 Drosophill
c 16	53.2	12.8	1001	13	CNS0155H	AL105023 Drosophill
c 17	53.2	12.8	1101	13	CNS0106X	AL098595 Drosophill
c 18	53	12.8	1101	13	CNS016LI	AL108896 Drosophill
c 19	52.6	12.7	734	13	CNS010MP	AL099163 Drosophill
c 20	52.6	12.7	1013	13	CNS00J7I	AL075824 Drosophill
c 21	52.6	12.7	1101	13	CNS00EPO	AL069493 Drosophill
c 22	52.4	12.6	524	13	CNS01U9O	AL167541 Tetraodon
c 23	52.4	12.6	996	13	CNS00FUH	AL071063 Drosophill
c 24	52.4	12.6	639	13	CNS038CX	AL232458 Tetraodon
c 25	51.8	12.5	928	13	CNS00DKY	AL071865 Drosophill
c 26	51.8	12.5	1101	13	CNS0039G	AL063921 Drosophill
c 27	51.8	12.5	1101	13	CNS00FMC	AL070972 Drosophill
c 28	51.6	12.4	992	13	CNS0562R	AL322812 Tetraodon
c 29	51.6	12.4	1043	13	CNS0145P	AL103735 Drosophill
c 30	51.6	12.4	1101	13	CNS00EO7	AL069440 Drosophill
c 31	51.4	12.4	876	13	CNS009G1	AL053529 Drosophill
c 32	51.2	12.3	678	13	CNS02A0C	AL187941 Tetraodon
c 33	51.2	12.3	1001	13	CNS0140O	AL103554 Drosophill
c 34	51.2	12.3	1101	13	CNS00EVL	AL069706 Drosophill
c 35	50.6	12.2	854	11	BF274512	BF274512 GA_ER002
c 36	50.6	12.2	963	10	AL566565	AL566565 Drosophill
c 37	50.6	12.2	1101	13	CNS00CYH	AL060100 Drosophill
c 38	50.6	12.2	1101	13	CNS0022U	AL097152 Drosophill
c 39	50.4	12.1	728	13	AQ272964	AQ272964 nbxb0028P
c 40	50.4	12.1	1101	13	CNS003BB	AL064089 Drosophill
c 41	50.4	12.1	1200	13	CNS016CO	AL106578 Drosophill
c 42	50.2	12.1	681	13	B60190	B60190 CIT-HSP-200
c 43	50	12.0	563	13	AQ326762	AQ326762 nbxb0038D
c 44	50	12.0	788	13	BH126604	BH126604 BARC-Satt
c 45	50	12.0	987	13	CNS014PQ	AL104456 Drosophill

ALIGNMENTS

RESULT 1

LOCUS CNS010MP

DEFINITION Drosophilla melanogaster genome survey sequence T7 end of BAC BACN04L20 of DrosBAC library from Drosophilla melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL099163

VERSION AL099163.1 GI:5610774

KEYWORDS GSS:

SOURCE fruit fly.

ORGANISM Plasmid Drosophilla melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophilla.

AUTHORS 1 (bases 1 to 734)

TITLE Genoscope.

JOURNAL Direct Submission

COMMENT Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophilla Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - . This Drosophilla melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES

source

1. 734

/organism="Drosophilla melanogaster"

/plasmid="pBelOBAC11"

/db\_xref="taxon:7227"

/clone\_lib="DrosBAC"

/clone="BACN04L20"





```
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db_xref="taxon:7227"  
/clone_lib="DrosBAC"  
/clone="BACN01A24"  
/note="end : sp6"
```

BASE COUNT	308 a	152 c	162 g	386 t	93 others
ORIGIN					

Query Match 13.8%; Score 57.2; DB 13; Length 1101;  
Best Local Similarity 41.4%; Pred. NO. 0.094;  
Matches 87; Conservative 38; Mismatches 85; Indels 0; Gaps 0;

[illegible][illegible]

ACCESSION AL106896  
VERSION AL106896.1 GI:5624374

GSS.  
fruit fly.  
Plasmid *Drosophila melanogaster*  
Eukaryota; Metazoa; Arthropoda;  
Pterygota; Neoptera; Endopterygota; Insecta;  
Muscomorpha; Ephydroidea; Diptera; brachycera;  
1 (bases 1 to 1101)  
Drosophilidae; *Drosophila*.  
Genoscope.  
AUTHORS

**TITLE**  
Journal

**COMMENT**  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqrefgenoscope.cns.fr](mailto:seqrefgenoscope.cns.fr))  
Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk>. This Drosophila melanogaster BAC  
library (Dros-BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBeloBAC11.

```

FEATURES
  source
    Location/Qualifiers
      1..1101
        /organism="Drosophila melanogaster"

```

BASE COUNT	203 a	220 c	84 g	158 t	436 others
RIGIN					
/note="end : t7"					

Query Match	13.3%	Score 55.4;	DB 13;	Length 1101;
Best Local Similarity	30.7%	Pred. No. 0.21;		

	Matches	96;	Conservative	79;	Mismatches	138;	Indels	0;	Gaps
QY	43	attacatattgaaactcttacggatgcagagaacaactcaccaagcattaatactgttcata	102	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
Dd	1050	AWMAAAWAAWATTTATTTHMTAATHTATWTWCATAAHWMWTTTTATATWAATH	991	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
QY	103	aatatgtcacattacgtatcatatacacgcgtatacaaatagtagcgagaataccatgt	162	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
Dd	990	TATTWATAHTATYTWTWHHWATWTWTWTATWANAHAANWASTAAAAYHHAATAIW	931	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
QY	163	aaagcagcaggggccaccatgdtctcaagtattataataattataattatcggtag	222	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
Dd	930	AANWATAAATTTWTWTWTATWTWTATTAATATWATWAAATAATWWWNAHHWWTYAAMW	871	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
QY	223	gatgtcacatgccgataagaaaagccaatttgtcgatgttaattccocatcttgaaagaaa	282	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
Dd	870	WWWATATWTATATWATAATWAAATTTWTATTAATTTATATWTTWAAAAATATATWTTAAAAAANA	811	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
QY	283	tatagttcaaattatttattgtataaataaaccaagtcaggattattagtcaccaagcaaaaaa	342	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
Dd	810	THAAWANAATAATWAAAAAAAHMTAAANAWATAAAAVAAAAAANAANAANAANAHAH	751	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
QY	343	taaattttatgat	355	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					
Dd	750	ATMAATTTTTTWT	738	: : : :   :   :   : : :   :   :   : : :   :   :   : : :   :   :   : : :					

RESULT	7
AQ506817/c	
LOCUS	
DEFINITION	AQ506817 537 bp DNA GSS 29-APR-1999 RPCI-11-281J17.TV RPCI-11 Homo sapiens genomic clone RPCI-11-281J17
ACCESSION	, DNA sequence.
VERSION	AQ506817
KEYWORDS	AQ506817.1 GI:4711564
SOURCE	GSS.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 537)
	Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
	J.C.

TITLE	JOURNAL	COMMENT
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building		
Unpublished (1997)		
Other-GSSs: RPCI-11-281J17.TJ		
Contact: Shaying Zhao, William Nierman, Mark Adams		
Department of Eukaryotic Genomics		
The Institute for Genomic Research		
9712 Medical Center Dr., Rockville, MD 20850		
Tel: 301 838 0200		
Fax: 301 838 0208		
Email: hbestigr.org		
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/ordering">http://bacpac.med.buffalo.edu/ordering</a> ) or from Research Genet cs ( <a href="mailto:info@resgen.com">info@resgen.com</a> ). BAC end search page: <a href="http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html">http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html</a> . Seq primer: w7		
Class: BAC ends.		

FEATURES source

BASE COUNT	216 a	11 c	30 q	28
/organism="Homo sapiens"				
/db_xref="GDB:7607752"				
/db_xref="taxon:9606"				
/clone="RPC11-281J17"				
/clone.lib="RPC1-11"				
/sex="Male"				
/cell_type="Lymphocytes"				
/note="Vector: pBACE3.6;"				
RPC11 Human Male BAC 11				



BASE COUNT	395 a	120 c	103 g	334 t	149 others
ORIGIN	/note="end : TTT3"				
Query Match	13.2%; Score 54.8; DB 13; Length 1101;				
Best Local Similarity	39.8%; Pred. No. 0.27;				
Matches	146; Conservative 32; Mismatches 189; Indels 0; Gaps 0;				
QY	28	ttttgactctccaccattacattgaaactcttcaggatgagacaactcacaaagcatt	87		
DB	548	TTTTWAAATTTTATATAAAAAA	111		
QY	88	aatcaggtccatataatgtcacattatcgtatatatacacogtatacaaatagtag	147		
DB	608	TTTTTTTTTTTTTTTTTTTTTTTTTTTATNTTTTAAATTTTAAATAAATTTTAT	667		
QY	148	cgaagaatccatgtcaagcagcagggggccaccatggttcacagttataataattataa	207		
DB	668	TTAWNAAAWATTTAAAAA	111		
QY	208	ttataattcgttgagatgcatcggcgcgataaagaagcgaattttagatggttaattc	267		
DB	728	TATATATWTTTAAWAAATATATATAAAAAA	111		
QY	268	coactctgaaagaatacagtttaaatatttattgataaataacaagtcagggtattata	327		
DB	788	ATTATATATWNNWATATWTTTAAATTTTATWATTTAAATTTAAATTTAAATTTAAW	847		
QY	328	gtccagcgaacaaacataaattattgcatgcaagtttaattcagaataatttcaataact	387		
DB	848	ATAAAAAA	111		
QY	388	gattata 394			
DB	908	AAAAAA 914			
RESULT 10					
CNS00JG1/c					
LOCUS	CNS00JG1 980 bp DNA GSS 03-JUN-1999				
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR38J18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL076232.1 GI:4955810				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 980)				
TITLE	Genoscope.				
JOURNAL	Direct Submission				
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequences : BP 191 91006 EVRY cedex - FRANCE (E-mail : secre@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammiger in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .				



## ORIGIN

[illegible]

RESULT	14
AU087444/c	
LOCUS	
DEFINITION	AU087444 500 bp mRNA EST AU087444 Sugano Malaria cDNA library Plasmodium falciparum CDNA clone XPEn669, mRNA sequence.
ACCESSION	AU087444
VERSION	AU087444.1 GI:12389585
KEYWORDS	EST.
SOURCE	malaria parasite P. falciparum. Plasmodium falciparum
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 500)
REFERENCE	Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S. FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite Plasmodium falciparum
TITLE	Nucleic Acids Res. 29 (1), 70-71 (2001) 20574754
JOURNAL	Contact: Junichi Watanabe
MEDLINE	Institute of Medical Science
COMMENT	The University of Tokyo, Department of Parasitology 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan Tel.: 81-3-5449-5378 Fax: 81-3-5449-5410

Email: jwacanab@nagae.ims.u-tokyo.ac.jp  
 Sukuzi.Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,  
 S. Construction and characterization of a full length-enriched and  
 a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).  
 Location:Qualification

FEATURES	source
a 3'-end-enriched cDNA library Gene 200 (1)	
Location/Qualifiers	
1. 500	
/organism="Plasmodium falciparum"	
/strain="3D7"	
/db_xref="taxon:5833"	
/clone="XPFn4669"	
/clone_lib="Sugano Malaria CDNA 1	
/dev_stage="erythrocytic stage"	
203 a	36 c
BASE COUNT	24 g 27 t 1
ORIGIN	

```

Query Match      12.9%   Score 53.4;   DB 10;   Length 500;
Best Local Similarity 50.3%;   Pred. No. 0.57;
Matches 155;   Conservative 0;   Mismatches 152;   Indels 1;   Gaps 1;

QY  89  acatggttcataaataatgtacattatcgttatatacacagttacacaatatag 147
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

D<sub>b</sub> 499 ATAAATTTAAATAAATAAATAACATATAGATATATATATATATATATTATATGTG 440  
Q<sub>y</sub> 148 cgaagaatacccatgtaaagcgcaggggcccacatggttccaagtattatatcaattataa 207  
D<sub>b</sub> 439 TTAATATACACTATATTTTTATTCGGTATATGTGCCCTCAATATTTTTTATTAATATAT 380  
Q<sub>y</sub> 208 ttataattatggtagtgtaacatgccgatgaagaaaaggccaatttgcagatggttaattc 267  
D<sub>b</sub> 379 TTAGAAAAATAGCTTTATCTAAAAATATAAAGAANAATTTTATTTTATATATTTATACN 320  
Q<sub>y</sub> 268 ccactctgaagaatacagttttaaatatttatgtataaaaaaacacgaagtcgggtattata 327  
D<sub>b</sub> 319 AATATATTTTATATATGTATATTAATTTATATATACATCGATGAAGAAAATAAAAAATAA 260  
Q<sub>y</sub> 328 gtccaaagcaaaaacataaattattgtgcgaagtttaaattcagaataattttccaataact 387  
D<sub>b</sub> 259 ATATACATATATATATATATATATAATATATATTAGTAAAAAGGATTGTGCTATTATA 200  
Q<sub>y</sub> 388 gattatat 395  
D<sub>b</sub> 199 AATATTAT 192

RESULT 15

CNS0182A	970 bp	DNA	GSS	26-JUL-1999
LOCUS				
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BAC37C16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL108796			
VERSION	AL108796.1	GI:5629100		
KEYWORDS	GSS.			
SOURCE	fruit fly.			
ORGANISM	Plasmid Drosophila melanogaster			
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
AUTHORS	1 (bases 1 to 970)			
TITLE	Genoscope.			
JOURNAL	Direct Submission			
	Submitted (23-JUL-1999)	Genoscope - Centre National de Sequences ;		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
	- Web : www.genoscope.cns.fr)			
	Determination of this BAC-end sequence was carried out as part of a			
	collaboration with the European Drosophila Genome Project (EDGP) -			
	http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC			
	library (Dros BAC) was made by Alain Billaud at CGPH (Centre			
	d'Etude du Polymorphisme Humain) with funding provided by a MRC			
	project grant. The DNA was prepared from embryos by Alain Bucheton			
	and Genevieve Payan. It has been constructed in the vector			
	pBelORAC11.			
COMMENT				

**FEATURES**

```

location/Qualifiers
1. 1970
/organism="Drosophila melanogaster"
/plasmid="pBelOBAc11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37C16"
/note="end : 87"
458 a 107 g 124 t 200 others
BASE COUNT
ORIGIN

```

Query Match 12.9%; Score 53.4; DB 13; Length 970;  
Best Local Similarity 42.6%; Pred. No. 0.51;  
Matches 144; Conservative 13; Mismatches 181; Indels 0; Gaps 0;

[illegible]



```
Db 75 AAAAAAAAAACCCATAATAAATCNACAAAAATTTAAAAATAATATAAAAAATAAAA 134
Qy 159 atgtaaagcagcagggggcaccatgggttccaagtattataataattataattatg 218
   ||:|||||:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 ATAWAAAAAAATWTTTATWAAWAAAAATATAWATAWAAAAAAANTTNCCTAATTTT 194
Qy 219 gtaggatgtacatggccgataaagaaaaagggcaatttggtagatgttaattcccatcttgaaa 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 ATTATNATAAAATAAANCCTATAAAAAAAATAAAAAATAAAAAATAAAAAACAATATAAAAA 254
Qy 279 gaaatatagttttaaatatttattgataaaaatacaaaagtcagggtattatagtcacagcaaa 338
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 ATAAAAAAWAAATAAAAAAAATAAAWAAWAAAAAAWAAAAAAATAAAWAAATAAAAAATA 314
Qy 339 aacataaatttattgatgcaagtttaaatcagaaata 376
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 AAAATAAATATATAAAAAAAATTAATWAAWAAAAAA 352
```

Search completed: February 15, 2002, 18:04:03  
Job time: 20952 sec

**THIS PAGE BLANK (USP 10)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:03:43 ; Search time 9904.61 Seconds  
(without alignments)  
16.274 Million cell updates/sec

Title: US-09-698-903B-4  
Perfect score: 15  
Sequence: 1 ntcgastwtsgwgtt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST.\*
- 1: em\_estfun.\*
  - 2: em\_esthum.\*
  - 3: em\_estin.\*
  - 4: em\_estom.\*
  - 5: em\_estpl.\*
  - 6: em\_estba.\*
  - 7: em\_estro.\*
  - 8: em\_estov.\*
  - 9: em\_htc.\*
  - 10: gb\_est1.\*
  - 11: gb\_est2.\*
  - 12: gb\_htc.\*
  - 13: gb\_gss.\*
  - 14: em\_gss\_fun.\*
  - 15: em\_gss\_hum.\*
  - 16: em\_gss\_inv.\*
  - 17: em\_gss\_pln.\*
  - 18: em\_gss\_pro.\*
  - 19: em\_gss\_rod.\*
  - 20: em\_gss\_vrt.\*
  - 21: em\_gss\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	12.4	82.7	101	13	CNS04Q4L	AL302142 Tetraodon
2	12.4	82.7	203	11	BG199530	RG199530 RST18921
3	12.4	82.7	204	10	BB258732	BB258732 BB258732
c 4	12.4	82.7	210	10	BB411504	BB411504 BB411504
c 5	12.4	82.7	216	10	AL367282	AL367282 MCB413H05
c 6	12.4	82.7	221	10	AV427766	AV427766 AV427766
7	12.4	82.7	226	11	BG057582	BG057582 nah92d01.
c 8	12.4	82.7	247	10	AV129451	AV129451 AV129451
9	12.4	82.7	248	10	BB074324	BB074324 BB074324
10	12.4	82.7	252	10	AV331577	AV331577 AV331577
c 11	12.4	82.7	257	10	BE530468	BE530468 M77D22STM
c 12	12.4	82.7	288	13	AZ921129	AZ921129 1006024D0

13	12.4	82.7	291	10	BB191477	BB191477
c 14	12.4	82.7	297	10	AV069588	AV069588
c 15	12.4	82.7	298	11	BF460492	BF460492 UI-M-CG0P
c 16	12.4	82.7	301	10	AJ284234	AJ284234 4A3B-AAU-
c 17	12.4	82.7	302	11	F14285	F14285 ATTS207 Or
c 18	12.4	82.7	307	10	AV558749	AV558749 AV558749
c 19	12.4	82.7	307	10	BB255970	BB255970 BB255970
c 20	12.4	82.7	310	10	AV421222	AV421222 AV421222
c 21	12.4	82.7	316	10	BB250639	BB250639 BB250639
c 22	12.4	82.7	335	11	BF098143	BF098143 ES7428664
c 23	12.4	82.7	342	10	AV209024	AV209024 AV209024
c 24	12.4	82.7	343	10	AV392544	AV392544 AV392544
25	12.4	82.7	344	10	AW693454	AW693454 NF065A10S
26	12.4	82.7	345	11	D40588	D40588 RICS2663A R
c 27	12.4	82.7	349	10	AV534046	AV534046 AV534046
c 28	12.4	82.7	355	10	AJ284258	AJ284258 4A3B-AAU-
c 29	12.4	82.7	355	10	AV441349	AV441349 AV441349
c 30	12.4	82.7	356	11	BF590524	BF590524 7h36h04.x
c 31	12.4	82.7	359	10	AV410421	AV410421 AV410421
c 32	12.4	82.7	360	11	C71547	C71547 C71547 XuJ1
c 33	12.4	82.7	366	10	BE321781	BE321781 NF009F02I
c 34	12.4	82.7	374	11	BF778123	BF778123 NXST-077
c 35	12.4	82.7	376	10	AA754245	AA754245 97MJ0056
c 36	12.4	82.7	376	10	AV202387	AV202387 AV202387
c 37	12.4	82.7	376	10	AV202461	AV202461 AV202461
c 38	12.4	82.7	377	10	AV422965	AV422965 AV422965
c 39	12.4	82.7	382	10	AV409979	AV409979 AV409979
c 40	12.4	82.7	382	10	AW796228	AW796228 MR2-UM002
c 41	12.4	82.7	386	11	BI241776	BI241776 RE38940.5
42	12.4	82.7	394	10	BE192723	BE192723 EST333452
c 43	12.4	82.7	397	11	BF483146	BF483146 WHE2315_C
44	12.4	82.7	405	11	BG630990	BG630990 cc-esf1cL
45	12.4	82.7	406	10	AW859801	AW859801 QV4-CT036

ALIGNMENTS

RESULT	1
CNS04Q4L/c	
LOCUS	101 bp DNA GSS 24-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-ORI end of clone 128H18 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL302142
VERSION	AL302142.1 GI:8179769
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
AUTHORS	1 (bases 1 to 101) Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 101) Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 101) Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis
COMMENT	

```

Query Match      82.7%;   Score 12.4;   DB 11;   Length 203;
Best Local Similarity 71.4%;   Pred. No. 3e+03;
Matches 10;   Conservative 4;   Mismatches 0;   Indels 0;   Gaps 0;

y      2   tcgastwtsgwggt 15
      ||||:|:|:|
b      38   TCGACTTTGGAGTT 51

```



XhoI; Plants were grown in an aeroponic chamber for 14 days on nitrogen-rich medium followed by 3 days on N-free medium. RNA was extracted from root tips (1-3 cm). cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zapR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exsist helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

BASE COUNT 69 a 56 c 34 g 56 t 1 others  
ORIGIN  
Query Match 82.7%; Score 12.4; DB 10; Length 216;  
Best Local Similarity 71.4%; Pred. No. 3e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
Db 74 TCGAGTTTCGAGTT 61

RESULT 6  
AV427766/c  
LOCUS  
DEFINITION AV427766 221 bp mRNA EST 23-MAY-2000  
japonicus cDNA clone MM086e02\_r 5', mRNA sequence.  
ACCESSION AV427766  
VERSION AV427766.1 GI:7788042  
KEYWORDS EST.  
SOURCE Lotus japonicus.  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;  
Lotus.  
REFERENCE 1 (bases 1 to 221)  
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
TITLE Generation of 7137 non-redundant expressed sequence tags from a  
legume, Lotus japonicus  
JOURNAL DNA Res. 7 (2), 127-130 (2000)  
MEDLINE 20277479  
COMMENT Contact: Yasukazu Nakamura  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
FEATURES  
source  
1. .221  
/organism="Lotus japonicus"  
/db\_xref="taxon:34305"  
/clone\_lib="MM086e02\_r"  
/clone="MM086e02\_r"  
/dev\_stage="young plants (two-week old)"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI; isolate=Miyakojima MG-20"  
BASE COUNT 73 a 60 c 25 g 63 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 221;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
Db 49 TCGACTTTCGAGTT 36

RESULT 7  
BG057582

LOCUS BG057582 226 bp mRNA EST 25-JAN-2001  
DEFINITION nah92d01.x1 NCI\_CGAP\_HN17 Homo sapiens cDNA clone IMAGE:4258200 3',  
mRNA sequence.  
ACCESSION BG057582  
VERSION BG057582.1 GI:12523216  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 226)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
unknown library type  
Seq primer: -40UP from Gibco.

FEATURES  
source  
1. .226  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4258200"  
/clone\_lib="NCI\_CGAP\_HN17"  
/tissue\_type="normal epithelium"  
/lab\_host="DH10B"  
/note="Organ: nasopharynx; Vector: pAMP10; mRNA made from  
normal nasopharyngeal epithelium, cDNA made by oligo-dr  
priming. Non-directionally cloned into UDG sites.  
Size-selected on agarose gel, average insert size 500 bp.  
Primary library. cDNA Library Preparation: David B.  
Krizman, Ph.D. REFERENCE: Krizman et al. (1996) Cancer  
Research 56:5380-5383."  
BASE COUNT 63 a 59 c 43 g 61 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 11; Length 226;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
Db 2 TCGACTTTGCTGTT 15  
RESULT 8  
AV129451/c  
LOCUS AV129451 247 bp mRNA EST 01-JUL-1999  
DEFINITION AV129451 Mus musculus C57BL/6J 11-day embryo Mus musculus CDNA  
clone 2700068H03, mRNA sequence.  
ACCESSION AV129451  
VERSION AV129451.1 GI:5315686  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 247)  
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Alzawa,K.,  
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara  
A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,  
Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H., Owa,C.,  
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara  
Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y., Tomimaga,N.,  
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,  
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

RIKEN  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
TITLE RIKEN Mouse ESTs  
JOURNAL Unpublished (1999)  
COMMENT Contact: Chie Owa  
Genome Science Laboratory  
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9445  
Fax: 81-298-36-9098  
Email: genome-res@rtc.riken.go.jp

Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
Source  
Location/Qualifiers  
1..247  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="2700068H03"  
/clone\_lib="Mus musculus C57BL/6J 11-day embryo"  
/sex="mixed"  
/dev\_stage="11-day embryo"  
77 a 41 c 34 g 95 t

Query Match 82.7%; Score 12.4; DB 10; Length 247;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
||||:|:|:|:|  
Db 121 TCGAGTTTGGTGT 108

RESULT 9  
BB074324 248 bp mRNA EST 27-JUN-2000  
LOCUS BB074324 RIKEN full-length enriched, adult male diencephalon Mus musculus cDNA clone 9330013E13 3' similar to S78234 nuc2 homolog, mRNA sequence.  
ACCESSION BB074324  
VERSION BB074324.1 GI:8584322  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 248)  
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigenoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Sueni-ro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

EMAIL: genome-res@rtc.riken.go.jp,  
URL:<http://genome.gsc.riken.go.jp/>  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermotabilization and thermoactivation of thermostable enzymes by

trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsuina,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
Source  
Location/Qualifiers  
1..248  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="9330013E13"  
/clone\_lib="RIKEN full-length enriched, adult male diencephalon"  
/sex="male"  
/tissue\_type="dienecephalon"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot - 10.0 and subtraction to Rot - 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATTAATATCCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 68 a 31 c 50 g 99 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 248;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 tcgastwtsgwgtt 15  
||||:|:|:|:|  
Db 197 TCGAGTATGTTGT 210

RESULT 10  
AV331577 252 bp mRNA EST 11-NOV-1999  
LOCUS AV331577 RIKEN full-length enriched, adult male medulla oblongata Mus musculus cDNA clone 6330524M18 3', mRNA sequence.  
DEFINITION AV331577  
ACCESSION AV331577  
VERSION AV331577  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 252)  
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,

RESULT	11
BE530468/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
COMMENT	

ALZ921129/C	LOCUS	DEFINITION
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

## STATISTICS

## REFERENCE

## REFERENCE

## REFERENCE

## REFERENCE



**TITLE**  
**JOURNAL**  
**COMMENT**

Maize genomic sequences found using engineered RescueMu transposon  
 Unpublished (2001)  
 Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Possible ligation site of ends cut by 2 different endonucleases.  
 Reverse complemented post-ligation sequence from source sequence.  
 Plate: 1006024 row: D column: 02  
 Class: transposon-tagged.  
 Location/Qualifiers

**FEATURES**

Source

1. .288  
 /organism="Zea mays"  
 /cultivar="mixed background W23/A188/B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="1006 - RescueMu Grid G"  
 /tissue\_type="leaf"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: Leaf; Vector: RescueMu (engineered from pluescript backbone); Site1: BamHI; Site2: BglII; RescueMu is a 4.9 Kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.znadb.iastate.edu' and follow the links for 'RescueMu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 74 a 91 c 73 g 50 t

Query Match 82.7%; Score 12.4; DB 13; Length 288;  
 Best Local Similarity 71.4%; Pred. No. 3.3e+03;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 133 TCGACTATCGAGTT 120

**RESULT 13**

BB191477

LOCUS

DEFINITION BB191477 291 bp mRNA EST 30-JUN-2000  
 musculus cDNA clone A330067M09 3' similar to AK001545 Homo sapiens  
 CDNA FLJ10683 fis. clone NT2RP3000142, highly similar to Homo  
 sapiens mRNA for KIAA0592 protein, mRNA sequence.

**ACCESSION**

BB191477

VERSION BB191477.1 GI:8852096

KEYWORDS

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 291)

Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci

P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,

Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,

Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,

Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.

Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.

Oono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata

Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sojabe,Y., Sugahara,Y.

Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya

T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamana,I.,

Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino

**TITLE**

**JOURNAL**

**COMMENT**

M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al.)  
 Unpublished (2000)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/  
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Thermostabilization and thermoactivation of thermolabile enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
 Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)

Please visit our web site (http://genome.rtc.riken.go.jp) for  
 further details.

**FEATURES**

source

Location/Qualifiers

1. .291  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="A330067M09"  
 /clone\_lib="RIKEN full-length enriched, adult male spinal  
 cord"  
 /sex="male"  
 /tissue\_type="spinal cord"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Site.1: SalI; Site.2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. cDNA went through one round of normalization  
 to Rot = 10.0 and subtraction to Rot = 459.0. Second  
 strand cDNA was prepared with the primer adapter of  
 sequence [5' GAGAGAGATTCGAGTTAATAAATTAATCCCCCCCCCCC  
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
 modified pBluescript KS(+) after bulk excision from Lambda  
 FLC I."

BASE COUNT 77 a 76 c 52 g 86 t

**ORIGIN**

Query Match 82.7%; Score 12.4; DB 10; Length 291;  
 Best Local Similarity 71.4%; Pred. No. 3.3e+03;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 179 TCGACTTTGGTGT 192

**RESULT 14**

AV069588/c

LOCUS

DEFINITION AV069588 297 bp mRNA EST 24-JUN-1999  
 Mus musculus small intestine C57BL/6J adult Mus musculus  
 cDNA clone 201031011, mRNA sequence.

```

ACCESSION AV069588
VERSION AV069588.1 GI:5189416
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 297)
Akihira, S., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara
, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tominaga, N.,
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
source
1..297
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="2010310L1"
/clone_lib="Mus musculus small intestine C57BL/6J adult"
/sex="male"
/tissue_type="small intestine"
/dev_stage="adult"

BASE COUNT 95 a 89 c 68 g 45 t
ORIGIN
|||||:|||||
2 tcgastwtsgwgtt 15
Db 74 TCGAGTTTGGTGTT 61

Query Match 82.7%; Score 12.4; DB 10; Length 297;
Best Local Similarity 71.4%; Pred. No. 3.3e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15
|||||:|||||
Db 74 TCGAGTTTGGTGTT 61

RESULT 15
BF460492/c
LOCUS BF460492 298 bp mRNA EST 04-DEC-2000
DEFINITION UI-M-CG0p-bmd-c-02-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
ACCESSION BF460492
VERSION BF460492
KEYWORDS EST.
SOURCE BF460492.1 GI:11529649
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 298)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 9704447
MEDLINE

COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
retina tissue cDNA library prepared at M.B. Soares Lab Clone
distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
GENETICS. It should be noted that Bento Soares is generating a
small number of additional specialized non-redundant arrays of BMAP
cDNAs whose availability will be considered under appropriate and
limited collaborative arrangements
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
1..298
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT7T30-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP Ret4_S2 library is a subtracted library,
ultimately derived from mouse retina tissue libraries at
various stages of development. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_L1B=NIH_BMAP_Ret4_S2
TAG_TISSUE=adult-retina
TAG_SEQ=GTGAGCGCGAC"

BASE COUNT 73 a 74 c 77 g 74 t
ORIGIN
|||||:|||||
2 tcgastwtsgwgtt 15
Db 121 TCGAGTTTGGAGTT 108

Query Match 82.7%; Score 12.4; DB 11; Length 298;
Best Local Similarity 71.4%; Pred. No. 3.3e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15
|||||:|||||
Db 121 TCGAGTTTGGAGTT 108

Search completed: February 15, 2002, 18:03:47
Job time: 20936 sec

```



**THIS PAGE BLANK**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:47:01 ; Search time 2553.1 Seconds  
(without alignments)  
161.541 Million cell updates/sec

Title: US-09-698-903B-5

Perfect score: 25  
Sequence: 1 ggatccccgatgagctaaagctagc 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_on.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	25	100.0	25	6	AX127752	AX127752 Sequence
2	25	100.0	25	6	AX172446	AX172446 Sequence
3	25	100.0	249	12	ARGMT0B	X05579 Soybean bet
4	25	100.0	1037	6	Al0942	Al0942 Nucleotide
5	25	100.0	1085	6	Al0939	Al0939 Nucleotide
6	25	100.0	1160	6	Al0943	Al0943 Nucleotide
c 7	25	100.0	2476	12	TBI251013	AJ251013 Transform
c 8	25	100.0	3236	12	TBI251014	AJ251014 Transform
c 9	25	100.0	4832	6	AX172441	AX172441 Sequence
c 10	25	100.0	4946	6	A60108	A60108 Sequence 1
c 11	25	100.0	4946	6	A76915	A76915 Sequence 1
c 12	25	100.0	4946	6	AR098307	AR098307 Sequence
c 13	25	100.0	4946	6	AX172440	AX172440 Sequence
c 14	25	100.0	5349	6	A71437	A71437 Sequence 7
c 15	25	100.0	5560	6	A60112	A60112 Sequence 5
c 16	25	100.0	5560	6	AR098311	AR098311 Sequence
c 17	25	100.0	5865	6	AX127748	AX127748 Sequence
c 18	25	100.0	5865	6	AX127748	AX127748 Sequence
c 19	25	100.0	6539	6	E31991	E31991 Mutated bar
c 20	25	100.0	6548	6	A60109	A60109 Sequence 2
c 21	25	100.0	6548	6	A76916	A76916 Sequence 2
c 22	25	100.0	6548	6	AR098308	AR098308 Sequence
c 23	25	100.0	6548	6	E31990	E31990 Mutated bar
c 24	25	100.0	7599	6	AX063413	AX063413 Sequence
c 25	25	100.0	12095	12	BINHYGDNA	Z37515 Binary vect
c 26	24	96.0	1166	6	A10941	A10941 Nucleotide
c 27	24	96.0	3200	6	I44104	I44104 Sequence 23
c 28	24	96.0	3201	6	I44103	I44103 Sequence 22
c 29	24	96.0	7566	6	A24783	A24783 plasmid pps
c 30	24	96.0	7566	6	AR074388	AR074388 Sequence
c 31	24	96.0	7639	6	A24782	A24782 plasmid pJD
c 32	24	96.0	7639	6	AR074387	AR074387 Sequence
c 33	24	96.0	7811	6	AR078675	AR078675 Sequence
c 34	19.2	76.8	134832	2	AP004068	AP004068 Oryza sat
c 35	18.8	75.2	152883	8	AC084319	AC084319 Oryza sat
c 36	18.6	74.4	175826	2	AC021890	AC021890 Homo sapi
c 37	17.8	71.2	25514	2	AC020475	AC020475 Drosophil
c 38	17.8	71.2	175118	3	AC010842	AC010842 Drosophil
c 39	17.8	71.2	188272	3	AC005639	AC005639 Drosophil
c 40	17.8	71.2	298616	3	AE003461	AE003461 Drosophil
c 41	17.6	70.4	83078	8	AF072897	AF072897 Arabidops
c 42	17.6	70.4	93489	8	AC005359	AC005359 REVERSE-C
c 43	17.6	70.4	149172	9	AC004826	AC004826 Homo sapi
c 44	17.6	70.4	149330	9	AL136084	AL136084 Human DNA
c 45	17.6	70.4	160570	2	AC092317	AC092317 Homo sapi

ALIGNMENTS

RESULT 1

AX127752 25 bp DNA PAT 15-MAY-2001  
LOCUS Sequence 5 from Patent WO0131042.  
DEFINITION AX127752  
ACCESSION AX127752  
VERSION AX127752.1 GI:14134399  
KEYWORDS . synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
1 (bases 1 to 25)  
REFERENCE Weston,B. and de Beuckeleer,M.  
AUTHORS Male-sterile brassica plants and methods for producing same  
TITLE Patent: WO 0131042-A 5 03-MAY-2001;  
JOURNAL Aventis CropScience N.V. (BE)  
FEATURES Location/Qualifiers  
source i. .25  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer MDB251"

BASE COUNT 6 a 8 c 7 g 4 t  
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaagctagc 25  
|||||  
Db 1 GGATCCCCCGATGAGCTAAGCTAGC 25

RESULT 2  
AXI72446  
LOCUS AXI72446 25 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 7 from Patent WO0141558.  
ACCESSION AXI72446  
VERSION AXI72446.1 GI:14597558  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequence.

REFERENCE 1 (bases 1 to 25)  
AUTHORS de Both, G. and de Beuckeleer, M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 7 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES Location/Qualifiers  
source 1..25  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer 251"

BASE COUNT 6 a 8 c 7 g 4 t  
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaagctagc 25  
|||||  
Db 1 GGATCCCCCGATGAGCTAAGCTAGC 25

RESULT 3  
ARGMTUB 249 bp DNA SYN 02-APR-1988  
LOCUS Soybean beta-1-tubulin gene fused to Ti plasmid unit 7 3UTR.  
DEFINITION X05579  
ACCESSION X05579.1 GI:58087  
VERSION beta-tubulin; fusion gene; plasmid.  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 249)  
AUTHORS Guiltinan, M.J., Velten, J., Bustos, M.M., Cyr, R.J., Schell, J. and Fosket, D.E.  
TITLE The expression of a chimeric soybean beta-tubulin gene in tobacco  
JOURNAL Mol. Gen. Genet. 207, 328-334 (1987)  
FEATURES Location/Qualifiers  
source 1..249  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

CDS 1..51  
/note="fusion product (17AA); Protein sequence is in  
Conflict with the conceptual translation"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA29084.1"  
/db\_xref="GI:4376141"  
/translation="AMAAWSSNSWSTDPMS"

misc\_feature 1..13  
/note="beta-1-tubulin sequence"  
misc\_feature 14..28  
/note="pUC 13 polylinker"  
misc\_feature 29..35  
/note="Sal I linker"  
misc\_feature 37

/note="theroretical fusion junction (24) with gene 7 of Ti  
plasmid"  
misc\_feature 173..178  
/note="put. polyA signal"  
polyA\_site 198  
/note="polyA site"  
misc\_feature 220..225  
/note="put. polyA signal"  
BASE COUNT 76 a 47 c 38 g 88 t  
ORIGIN

Query Match 100.0%; Score 25; DB 12; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.043;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaagctagc 25  
|||||  
Db 36 GGATCCCCCGATGAGCTAAGCTAGC 60

RESULT 4  
AL0942  
LOCUS AL0942 1037 bp DNA PAT 27-SEP-1993  
DEFINITION Nucleotide sequence 4 from patent number DE3920034.  
ACCESSION AL0942  
VERSION AL0942.1 GI:492369  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified.

REFERENCE 1 (bases 1 to 1037)  
AUTHORS  
JOURNAL Patent: DE 3920034-A 4 31-MAY-1990;  
FEATURES Location/Qualifiers  
source 1..1037  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 338 a 174 c 166 g 359 t  
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1037;  
Best Local Similarity 100.0%; Pred. No. 0.039;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaagctagc 25  
|||||  
Db 767 GGATCCCCCGATGAGCTAAGCTAGC 791

RESULT 5  
AL0939  
LOCUS AL0939 1085 bp DNA PAT 27-SEP-1993  
DEFINITION Nucleotide sequence 1 from patent number DE3920034.  
ACCESSION AL0939  
VERSION AL0939.1 GI:492367  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified.

REFERENCE 1 (bases 1 to 1085)  
AUTHORS  
JOURNAL Patent: DE 3920034-A 1 31-MAY-1990;  
FEATURES Location/Qualifiers  
source 1..1085

```
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      369 a   218 c   155 g   343 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 1085;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatccccgatgagctaaagctagc 25
|||||
Db 815 GGATCCCCCGATGAGCTAAGCTAGC 839

RESULT 6
LOCUS      A10943      1160 bp      DNA      PAT      27-SEP-1993
DEFINITION Nucleotide sequence 5 from patent number DE3920034.
ACCESSION  A10943
VERSION     A10943.1 GI:492370
KEYWORDS   .
SOURCE     unidentified.
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 1160)
AUTHORS   .
JOURNAL   Patent: DE 3920034-A 5 31-MAY-1990;
FEATURES   Location/Qualifiers
            source
            1..1160
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT      367 a   194 c   188 g   411 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 1160;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatccccgatgagctaaagctagc 25
|||||
Db 890 GGATCCCCCGATGAGCTAAGCTAGC 914

RESULT 7
TBI251013/c
LOCUS      TBI251013      2476 bp      DNA      SYN      26-JUN-2001
DEFINITION Transformation binary vector pBAR-A, T-DNA region.
ACCESSION  AJ251013
VERSION     AJ251013.1 GI:6453666
KEYWORDS   t-dna; transformation binary vector.
SOURCE     Transformation binary vector pBAR-A.
ORGANISM   artificial sequence; vectors.
REFERENCE  1 (bases 1 to 2476)
AUTHORS   Cardon,G.H. and Huijser,P.
TITLE     T-DNA sequence of the plant transformation binary vector pBAR-A
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 2476)
AUTHORS   Cardon,G.H.
TITLE     Direct Submission
JOURNAL   Submitted (12-NOV-1999) Cardon G.H., Molecular Plant Genetics,
          Max-Planck- Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg
          10, 50829 Cologne, GERMANY
COMMENT   Related sequences: U99365, J01825, J01636, X02513, V00090, X05822,
          X01077, J01826.
FEATURES   Location/Qualifiers
            source
            1..2476
            /organism="Transformation binary vector pBAR-A"
            /db_xref="taxon:108140"
            /note="Can be cultured in Escherichia coli or
            Agrobacterium tumefaciens-T-DNA region of a binary vector

derived from plasmid pGPTV-BAR"
1..148
/note="similar to T-DNA left border"
complement(149..279)
/gene="laci"
CDS
complement(<149..>279)
/gene="laci"
/codon_start=1
/transl_table=11
/protein_id="CAC42834.1"
/db_xref="GI:14572627"
/translation="NHQTGFSPAGANQRGLATLSGPGGEGQSAVARLTGEKKNHP"

misc_feature
gene
CDS
rep_origin
280..576
/note="M13 ori"
/direction=RIGHT
complement(<577..>732)
/gene="lacZ"
gene
complement(577..732)
/gene="lacZ"
733..955
/gene="transcript 7"
733..955
/gene="transcript 7"
mRNA
complement(<956..>1544)
/gene="bar"
gene
complement(956..1544)
/gene="bar"
1545..1837
/gene="NOS"
1545..2314
/gene="NOS"
1838..2157
/gene="NOS"
/note="similar to transposable element Tn5"
2158..2191
/gene="NOS"
/note="synthetic multiple cloning site"
2192..2314
/gene="NOS"
2315..2476
/note="similar to T-DNA right border"
BASE COUNT      563 a   614 c   699 g   600 t
ORIGIN

Query Match      100.0%; Score 25; DB 12; Length 2476;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatccccgatgagctaaagctagc 25
|||||
Db 961 GGATCCCCCGATGAGCTAAGCTAGC 937

RESULT 8
TBI251014/c
LOCUS      TBI251014      3236 bp      DNA      SYN      26-JUN-2001
DEFINITION Transformation binary vector pBAR-35S, T-DNA region.
ACCESSION  AJ251014
VERSION     AJ251014.1 GI:6453667
KEYWORDS   t-dna; transformation binary vector.
SOURCE     Transformation binary vector pBAR-35S.
ORGANISM   Transformation binary vector pBAR-35S.
REFERENCE  1 (bases 1 to 3236)
AUTHORS   Cardon,G.H. and Huijser,P.
TITLE     T-DNA sequence of the plant transformation binary vector pBAR-A
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 3236)
AUTHORS   Cardon,G.H.
TITLE     Direct Submission
JOURNAL   Submitted (12-NOV-1999) Cardon G.H., Molecular Plant Genetics,
```

Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg  
10, 50829 Cologne, GERMANY  
Related sequences: U09365, J01825, J01636, X02513, V00090, X05822  
X01077, J01826.

# FEATURES

```

source
1..3236
    /organism="Transformation binary vector pBAR-35s"
    /db_xref="taxon:108141"
    /note="Can be cultured in Escherichia coli or
    Agrobacterium tumefaciens-T-DNA region of a binary vector
    derived from plasmid pGPTV-BAR"
1..148
    /note="similar to T-DNA left border"
    /complement(149..279)
    /gene="lacI"
    /complement(<149..>279)
    /gene="lacI"
    /codon_start=1
    /transl_table=11
    /protein_id="CAC42835.1"
    /db_xref="GI:14572628"
    /translation="NHHQTGFSPAGANQRPLAATLSPGGGQSAVARLTGEKKNHP"

misc_feature
280..576
    /note="M13 ori"
    /complement(<577..>732)
    /gene="lacZ"
    /complement(577..732)
    /gene="lacZ"
733..955
    /gene="transcript 7"
733..955
    /gene="transcript 7"
    /note="putative"
    /complement(<956..>1544)
    /gene="bar"
    /complement(956..1544)
    /gene="bar"
1545..1837
    /gene="NOS"
1545..1837
    /gene="NOS"
1838..2157
    /note="similar to transposable element Tn5"
2167..2707
    /gene="camV 35S"
2167..2956
    /gene="camV 35S"
2708..2730
    /gene="camV 35S"
    /note="Synthetic multiple cloning site"
2731..2956
    /gene="camV 35S"
2957..3074
    /gene="NOS gene promoter"
    /gene="NOS gene promoter"
3075..3236
    /note="similar to T-DNA right border"

BASE COUNT      804 a  786 c  860 g  786 t
ORIGIN

Query Match      100.0%; Score 25; DB 12; Length 3236;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25
    |||
Db 961 GGATCCCCGATGAGCTAAGCTAGC 937

RESULT 9

```

## AX172441/c

```

LOCUS      AX172441      4832 bp      DNA
DEFINITION Sequence 2 from Patent WO0141558.
ACCESSION  AX172441
VERSION     AX172441.1 GI:14597553
KEYWORDS    synthetic construct.
SOURCE      synthetic construct
            artificial construct
REFERENCE   1 (bases 1 to 4832)
AUTHORS    de Both, G. and de Beuckeleer, M.
TITLE       Hybrid winter oilseed rape and methods for producing same
JOURNAL     Patent: WO 0141558-A 2 14-JUN-2001;
            Aventis CropScience N.V. (BE)
FEATURES    Location/Qualifiers
             1..4832
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="T-DNA of plasmid pTHW118"
             misc_feature
             1883..4065
                /note="HpaI restriction fragment"
BASE COUNT  1528 a   883 c   932 g  1488 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 6; Length 4832;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ggatcccccgatgagctaaagctagc 25
    |||
Db 317 GGATCCCCGATGAGCTAAGCTAGC 293

```

## RESULT 10

```

A60108/c
LOCUS      A60108      4946 bp      DNA
DEFINITION Sequence 1 from Patent WO9706267.
ACCESSION  A60108
VERSION     A60108.1 GI:3715124
KEYWORDS    Transformation vector pTHW107.
SOURCE      Transformation vector pTHW107
            artificial sequence; vectors.
ORGANISM    De,B.M.
REFERENCE   1 (bases 1 to 4946)
AUTHORS     Genetic transformation using a PARP inhibitor
            Patent: WO 9706267-A 1 20-FEB-1997;
            PLANT GENETIC SYSTEMS NV (BE)
JOURNAL     Location/Qualifiers
FEATURES    1..4946
            /organism="Transformation vector pTHW107"
            /db_xref="taxon:126810"
BASE COUNT  1569 a   891 c   963 g  1523 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 6; Length 4946;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ggatcccccgatgagctaaagctagc 25
    |||
Db 317 GGATCCCCGATGAGCTAAGCTAGC 293

```

## RESULT 11

```

A76915/c
LOCUS      A76915      4946 bp      DNA
DEFINITION Sequence 1 from Patent EP0757102.
ACCESSION  A76915
VERSION     A76915.1 GI:6088712
KEYWORDS

```



```

/db_xref="taxon:32630"
/note="r-DNA of plasmid pTHW107"
964..4906
misc_feature
/note="Hind III fragment"
1569 a 891 c 963 g 1523 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 6;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0;

Qy 1 ggatcccccgatgagctaagctagc 25
|||||
Db 317 GGATCCCCGGATGAGCTAAGCTAGC 293
|||||

```

RESULT 14  
A71437/C

LOCUS A71437 5349 bp DNA  
DEFINITION Sequence 7 from Patent WO9810081.  
ACCESSION A71437  
VERSION A71437.1 GI:4775050

SOURCE: ORGANISM  
unclassified.  
unclassified.  
unclassified.  
1 (bases 1 to 5349)  
Michiels, F. and Williams, M.  
IMPROVED BARSTAR GENE  
Patent: WO 9810081-A 7 12-MAR-1998;  
MICHELIS FRANK (BE)

```

BASE COUNT      1339 a  1233 c  1290 g  1487 t
ORIGIN
1. .5349
/organism="unidentified"
/db_xref="taxon:32644"
location/Qualifiers
source

```

Query Match	100.0%	Score 25	DB 6	Length 5349
Best Local Similarity	100.0%	Pred. No. 0.035		
Matches 25	Conservative	0	Mismatches 0	Indels 0
Gaps	0			

Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT	15
A60112/c	
LOCUS	A60112      5560 bp      DNA
DEFINITION	Sequence 5 from Patent WO9706267.

VERSION	AB0112.1	GI:13715120
KEYWORDS	Transformation vector pTHW142.	
SOURCE	Transformation vector pTHW142	
ORGANISM	artificial sequence; vectors.	
REFERENCE	1 (bases 1 to 5560)	

AUTHORS	DE, B. M.
TITLE	Genetic transformation using a PARP inhibitor
JOURNAL	Patent: WO 9706267-A 5 20-FEB-1997;
	PLANT GENETIC SYSTEMS NV (BE)
FEATURES	Location/Qualifiers
SOURCE	1. .5560 /organism="Transformation vector pTHW142" /db_xref="taxon:125817"
BASE COUNT	1533 a 1199 c 1244 g 1576 t
ORIGIN	8 others
Query Match	100.0% Score 25: DB 6: Length 55

Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatccccgatgagctaagctagc 25  
Db 304 GGATCCCGCGATGAGCTAAGCTAGC 280

Search completed: February 15, 2002, 18:47:05  
Job time: 20064 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:01:14 ; Search time 868.33 Seconds  
(without alignments)  
24.683 Million cell updates/sec

Title: US-09-698-903B-5

Perfect score: 25  
Sequence: 1 ggatccccgatgactaagtagc 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	AAH25426	PCR primer for fla
2	25	100.0	25	AAD06994	PCR primer MDR251
3	25	100.0	1037	AAQ04705	USP-Promoter-casse
4	25	100.0	1085	AAQ04703	Legumin-signalpept
5	25	100.0	1160	AAQ04706	USP-signalpeptide
6	25	100.0	1303	AAAT39337	Plasmid pTS88 (Eco
7	25	100.0	4832	AAH25423	Nucleotide sequenc
8	25	100.0	4946	AAH25422	T-DNA of plasmid p
9	25	100.0	4946	AAH25422	Nucleotide sequenc
10	25	100.0	5228	AAF86439	Plasmid pTS172del
11	25	100.0	5349	AAV23239	T-DNA of pTTS24.

12	25	100.0	5864	17	AAT39339	Plasmid pTCO113 T-
13	25	100.0	5864	17	AAT39339	Plasmid pTCO113 T-
14	25	100.0	5865	22	AAD06990	Chimeric T-DNA of
15	25	100.0	5865	22	AAD06990	Chimeric T-DNA of
16	25	100.0	6539	21	AAZ91097	E. coli plasmid pT
17	25	100.0	6548	17	AAT39336	Plasmid pTS174 use
18	25	100.0	6548	18	AAZ91096	Plasmid pTS172. C
19	25	100.0	6548	21	AAZ91096	E. coli plasmid pT
20	25	100.0	7492	22	AAF86441	Plasmid pTS346. U
21	25	100.0	7599	22	AAZ25320	Nucleotide sequenc
22	24	96.0	1166	11	AAQ04704	USP-Promoter-casse
23	24	96.0	3201	12	AAQ14529	pPS29 Bt ICP codi
24	24	96.0	3201	12	AAQ15144	pVE36 Bt ICP codin
25	24	96.0	7566	14	AAQ42160	Plasmid pPS0212 co
26	24	96.0	7639	14	AAQ42159	Plasmid pJD884 con
27	19	76.0	34	22	AAH43929	Agrobacterium tume
28	19	76.0	34	22	AAH25165	PCR primer used to
29	17.2	68.8	2562	22	AAI59154	Human polynucleoti
30	17.2	68.8	6380	22	AAI57822	Human polynucleoti
31	17.2	68.8	7104	22	AAI59608	Human polynucleoti
32	17	68.0	29	14	AAQ42148	Bt884 oligonucleot
33	17	68.0	3153	21	AAZ29122	Plasmid pDV131 comp
34	17	68.0	3336	21	AAZ29121	Plasmid pDV130 comp
35	17	68.0	3694	21	AAZ29124	Plasmid pDV133 used
36	17	68.0	3877	21	AAZ29123	Plasmid pDV132 used
37	17	68.0	24593	6	AAH50226	Sequence of opine
38	17	68.0	24596	6	AAH50182	Complete nucleotid
39	16.6	66.4	3488	19	AAV46307	S. pneumoniae deri
40	16.6	66.4	6204	19	AAV44592	Streptococcus pneu
41	16.6	66.4	534720	19	AAV30458	Rhizobium species
42	16.6	66.4	536165	19	AAV30459	Rhizobium species
43	16.2	64.8	457	21	AAZ53867	Arabidopsis thalia
44	16.2	64.8	521	21	AAZ54447	Arabidopsis thalia
45	16.2	64.8	526	21	AAZ52318	Arabidopsis thalia

ALIGNMENTS

RESULT 1  
AAH25426  
ID AAH25426 standard; DNA; 25 BP.  
XX AAH25426;  
XX AC  
XX AC  
XX 22-AUG-2001 (first entry)  
XX PCR primer for flanking regions in transgenic plant MS-BN1.  
XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
KW fertility restorer gene; barnase gene; PCR primer; ss.  
XX Synthetic.  
XX OS  
XX WO200141558-A1.  
XX 14-JUN-2001.  
XX 06-DEC-2000; 2000WO-EPI2872.  
XX 08-DEC-1999; 99US-0457037.  
XX (AVET ) AVENTIS CROPS SCIENCE NV.  
XX De Both G, De Beuckeleer M;  
XX WPI; 2001-381419/40.  
XX Transgenic winter oilseed rape plants suited for producing hybrid seed  
PT with improved qualities, comprises a male-sterility gene and fertility  
PT restorer gene, integrated into the genome -  
XX Example 4; Page 42; 98pp; English.  
PS

XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAH25425-26 were used to amplify the flanking region of  
 CC a vector in a transgenic plant which carries the TA29-barnase  
 CC transgene.

XX  
 SQ Sequence 25 BP; 6 A; 8 C; 7 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0053;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25  
 |||||  
 Db 1 ggatcccccgatgagctaaagctagc 25

## RESULT 2

AA06994  
 ID AAD06994 standard; DNA; 25 BP.

XX  
 AC AAD06994;

DT 06-AUG-2001 (first entry)

DE PCR primer MDS251 to generate the flanking region of elite event MS-B2.

KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; thermal asymmetric interlaced;  
 KW TAIL; ss.

XX Agrobacterium sp.

XX WO200131042-A2.

PD 03-MAY-2001.

XX 26-OCT-2000; 2000WO-EP10680.

XX 29-OCT-1999; 99US-0430497.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX Weston B, De Beuckeleer M;

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -

PS Example 3; Page 28; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.

CC The present sequence is primary thermal interlaced (TAIL)-PCR primer  
 CC MDS251 used to right (5') and left (3') border flanking region of elite  
 CC event MS-B2. This primer corresponds to position 293-317 of plasmid  
 CC pTC0113.

XX  
 SQ Sequence 25 BP; 6 A; 8 C; 7 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0053;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25  
 |||||  
 Db 1 ggatcccccgatgagctaaagctagc 25

## RESULT 3

AA04705  
 ID AA04705 standard; DNA; 1037 BP.

XX  
 AC AA04705;

DT 12-OCT-1990 (first entry)

XX USP-Promoter-cassette USP-Pr.T7.1.

DE Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.

XX DE3920034-A.

XX 31-MAY-1990.

XX 20-JUN-1989; 89DE-3920034.

XX 19-SEP-1988; 88DD-0319887.

XX (PFLA-) VE KOMB PFLANZENZUC.

XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

XX WPI; 1990-172459/23.

XX Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.

XX Disclosure; ; pp; German.

XX The unique BglII-Ort (720-725) site is for ligating foreign DNA and the  
 CC HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 CC cassette in the Ti-vector pGA471. Agrobacterium tumefaciens is  
 CC transfected.

XX See also AAQ04703-Q04706.

XX  
 SQ Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 100.0%; Score 25; DB 11; Length 1037;  
 Best Local Similarity 100.0%; Pred. No. 0.007;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25  
 |||||  
 Db 767 ggatcccccgatgagctaaagctagc 791

## RESULT 4

AA04703  
 ID AA04703 standard; DNA; 1085 BP.

XX  
 AC AA04703;

XX 12-OCT-1990 (first entry)

XX Legumin-signalpeptide cassette Le-Sig.T7.

XX

```
KW Foreign DNA incorporation; recombinant DNA techniques;
XX higher plant genome; legumin; B4 gene; signalpeptide; Le-Sig.T7.; ss.
FH Key Location/Qualifiers
FT CDS 747..814
FT /*tag= a
FT /product=legumin-signalpeptide
XX
XX DE3920034-A.
XX 31-MAY-1990.
XX
XX 20-JUN-1989; 89DE-3920034.
XX
XX 19-SEP-1988; 88DD-0319887.
XX
XX (PFLA-) VE KOMB PFLANZENZUC.
XX
XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;
XX
XX WPI; 1990-172459/23.
XX P-PSDB; AAR05198.
XX
XX Incorporation of DNA into higher plant genome - by specified
XX recombinant DNA techniques.
XX
XX Disclosure; ; pp; German.
XX
XX The Legumin gene B4 is used. The unique BglIII-Ort (815-820) site is for
XX ligating foreign DNA and the HindIII-Ort in the 3' polylinker
XX (1080-1085) for cloning the cassette in the Ti-vector pGA471.
XX The cassette is cloned into the binary Ti-vectors pGA471 and
XX Agrobacterium tumefaciens is transfected.
XX See also AAQ04703-Q04706.
XX
XX Sequence 1085 BP; 368 A; 219 C; 155 G; 343 T; 0 other;

Query Match 100.0%; Score 25; DB 11; Length 1085;
Best Local Similarity 100.0%; Pred. No. 0.007;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaagctagc 25
| | | | | | | | | | | | | | | | |
Db 815 ggatcccccgatgagctaagctagc 839

RESULT 5
AAQ04706
ID AAQ04706 standard; DNA; 1160 BP.
XX
XX AC AAQ04706;
XX
XX DT 12-OCT-1990 (first entry)
XX
XX USP-signalpeptide cassette USP-Sig.T7.
XX
XX Foreign DNA incorporation; recombinant DNA techniques;
XX higher plant genome; signalpeptide; USP-Sig.T7.; ss.
XX
XX Key Location/Qualifiers
XX CDS 708..877
XX /*tag= a
XX /product=signalpeptide
XX 747..817
XX
XX intron
XX
XX DE3920034-A.
XX
XX 31-MAY-1990.
XX
XX 20-JUN-1989; 89DE-3920034.
XX
XX 19-SEP-1988; 88DD-0319887.
XX
```

```
XX (PFLA-) VE KOMB PFLANZENZUC.
XX
XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;
XX
XX WPI; 1990-172459/23.
XX P-PSDB; AAR05199.
XX
XX Incorporation of DNA into higher plant genome - by specified
XX recombinant DNA techniques.
XX
XX Disclosure; ; pp; German.
XX
XX The unique BglIII-Ort (890-895) site is for
XX ligating foreign DNA and the HindIII-Ort in the 3' polylinker
XX (1155-1160) for cloning the cassette in the Ti-vector pGA471.
XX The cassette is cloned into the binary Ti-vectors pGA471 and
XX Agrobacterium tumefaciens is transfected.
XX See also AAQ04703-Q04706.
XX
XX Sequence 1160 BP; 369 A; 192 C; 188 G; 411 T; 0 other;

Query Match 100.0%; Score 25; DB 11; Length 1160;
Best Local Similarity 100.0%; Pred. No. 0.007;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaagctagc 25
| | | | | | | | | | | | | | | | |
Db 890 ggatcccccgatgagctaagctagc 914

RESULT 6
AAT39337
ID AAT39337 standard; DNA; 1303 BP.
XX
XX AC AAT39337;
XX
XX DT 22-JAN-1997 (first entry)
XX
XX DE Plasmid pTS88 (EcoRI-HindIII fragment).
XX
XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
XX transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX misc_feature 1..35
XX /*tag= a
XX /label= pGEM2
XX /note= "polylinker of pGEM2"
XX
XX promoter 36..694
XX /*tag= b
XX /label= p35S
XX /function= 35S promoter of cauliflower mosaic virus
XX strain CM1841
XX
XX CDS 695..967
XX /*tag= c
XX /label= barstar
XX /product= Bacillus amyloliquefaciens barstar
XX
XX polyA_signal 968..1287
XX /*tag= d
XX /label= 3'g7
XX /function= region containing polyadenylation signal
XX of gene 7 og Agrobacterium T-DNA
XX
XX misc_feature 1288..1303
XX /*tag= e
XX /label= pGEM2
XX /note= "polylinker of pGEM2"
XX
XX
XX WO9626283-A1.
XX
```

```

PD 29-AUG-1996.
XX
PF 21-FEB-1996; 96WO-EP00722.
XX
PR 21-FEB-1995; 95EP-0400364.
XX
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
PI Botterman J, Cornelissen M, Michiels F;
XX WPI; 1996-402373/40.
DR
XX
XX Prodn. of male sterile plants by transforming with a chimaeric
PT construct - comprising a male sterility DNA e.g. barnase and a
PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
PT for generating hybrid cultivars
XX
XX Example 1; Page 38; 56pp; English.
PS
XX The HindIII-EcoRI fragment (AAT39337) of plasmid pT888 contains
CC barstar DNA under control of a 35S promoter. The plasmid was
CC used with pT8174 (see also AAT39336) contg. barnase DNA under
CC control of the stamen-specific promoter EI to produce male sterile
CC rice cv. Kochihibiki transgenic plants, and with plasmid pVE136
CC (see also AAT39338) contg. barnase DNA under control of the stamen-
CC specific PCA55 promoter to produce male sterile maize plants.
CC Expression of barnase (a ribonuclease) in the stamen leads to male
CC sterility. Constitutive expression of barstar counteracts possible
CC low level expression of barnase DNA in non-stamen tissue.
XX
XX Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;
SQ
Query Match 100.0%; Score 25; DB 17; Length 1303;
Best Local Similarity 100.0%; Pred. No. 0.0071;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ggatcccccgatgagctaaagctagc 25
|||||
Db 1024 ggatcccccgatgagctaaagctagc 1048
|||||

RESULT 7
AAH25423/c
ID AAH25423 standard; DNA; 4832 BP.
XX
XX AAH25423;
XX
XX 22-AUG-2001 (first entry)
XX
XX Nucleotide sequence of plasmid pTHW118.
XX
XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barstar gene; ss.
XX
XX Synthetic.
OS Streptomyces hygroscopicus.
OS Arabidopsis thaliana.
OS Bacillus amyloliquefaciens.
OS Nicotiana tabacum.
XX
XX Key Location/Qualifiers
FH misc_feature 1..25
FT /*tag= a
FT /note= "right border repeat from TL-DNA from pTiB6S3"
FT 26..53
FT /*tag= b
FT /note= "synthetic polylinker derived sequences"
FT 54..90
FT /*tag= c
FT /note= "residual sequence from TL-DNA at right
FT border repeat"
FT 91..97

```

---

```

FT /*tag= d
FT /note= "synthetic polylinker derived sequences"
FT complement (98..309)
FT /*tag= e
FT /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"
FT 310..330
FT /*tag= f
FT /note= "synthetic polylinker derived sequences"
FT complement (331..882)
FT /*tag= g
FT /note= "Streptomyces hygroscopicus bialaphos
FT resistance (bar) gene"
FT complement (883..2608)
FT /*tag= h
FT /note= "atsLA ribulose-1,5-biphosphate carboxylase
FT small subunit gene from Arabidopsis thaliana"
FT 2609..2658
FT /*tag= i
FT /note= "synthetic polylinker derived sequences"
FT complement (2659..2919)
FT /*tag= j
FT /note= "TaqI fragment from 3' UTR of nopaline
FT synthase gene from T-DNA of pTiT37 and
FT containing plant polyadenylation signals"
FT 2920..2940
FT /*tag= k
FT /note= "Synthetic polylinker derived sequences"
FT 2941..2980
FT /*tag= l
FT /note= "downstream of Bacillus amyloliquefaciens
FT barstar coding region"
FT complement (2981..3253)
FT /*tag= m
FT /note= "Barstar gene coding region from Bacillus
FT amyloliquefaciens"
FT complement (3254..4762)
FT /*tag= n
FT /note= "another-specific gene TA29 promoter from
FT Nicotiana tabacum"
FT 4763..4807
FT /*tag= o
FT /note= "synthetic polylinker derived sequences"
FT 4808..4832
FT /*tag= p
FT /note= "left border repeat from TL-DNA from pTiB6S3"
XX
XX WO200141558-A1.
XX
XX 14-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-EP12872.
XX
XX 08-DEC-1999; 99US-0457037.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX De Both G, De Beuckeleer M;
XX WPI; 2001-381419/40.
XX
XX Transgenic winter oilseed rape plants suited for producing hybrid seed
PT with improved qualities, comprises a male-sterility gene and fertility
PT restorer gene, integrated into the genome -
XX
XX Example 1; Page 80-82; 98pp; English.
XX
XX The specification describes a pair of transgenic winter oilseed rape
CC plants suited for producing hybrid seed. One of the plants has an
CC expression cassette comprising a male-sterility gene, and the other
CC plant has an expression cassette comprising a fertility restorer gene,
CC integrated into the genome. The fertility restorer gene is capable of
CC preventing the activity of the male-sterility gene. The plant pair is
CC useful for producing hybrid seed. Plants developed from the hybrid

```

CC seed have agronomic performance, genetic stability and adaptability to  
 CC different genetic backgrounds. The present sequence represents  
 CC plasmid pTHW118. This plasmid comprises the barstar gene, which acts as  
 CC a fertility restorer gene. The plasmid is used to create transgenic  
 CC plants of the invention.

XX SQ Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;

Query Match 100.0%; Score 25; DB 22; Length 4832;  
 Best Local Similarity 100.0%; Pred. No. 0.0078;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgcgatgagctaagctagc 25  
 Db 317 GGATCCCGCGATGAGCTAAGCTAGC 293

RESULT 8  
 AAT59531/C  
 ID AAT59531 standard; DNA; 4946 BP.

XX AC AAT59531;

XX DT 07-MAY-1997 (first entry)

XX DE T-DNA of plasmid pTHW107.

XX KW Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PARP;  
 KW niacinamide; Agrobacterium; T-DNA; male sterile; barnase;  
 KW ribonuclease; Rnase; cereal; wheat; oilseed rape; Brassica napus;  
 KW plasmid pTHW107; ds.

XX OS Chimeric Agrobacterium sp.;  
 OS Chimeric Arabidopsis thaliana;  
 OS Chimeric Nicotiana tabacum.

FH Key Location/Qualifiers  
 FT misc\_RNA complement (1..25)

FT /\*tag= a

FT /label= RB

FT /note= "T-DNA right border"

FT /\*tag= b

FT /label= 3'g7

FT /note= "3' untranslated region contg. the poly-A  
 signal of gene-7 of Agrobacterium T-DNA"

FT complement (331..882)

FT /\*tag= c

FT /label= Bar

FT /product= phosphinothricin acetyltransferase

FT complement (883..2608)

FT /\*tag= d

FT /label= PSSU

FT /note= "promoter region of Rubisco small subunit  
 gene of Arabidopsis thaliana"

FT complement (2658..3031)

FT /\*tag= e

FT /label= 3'nos

FT /note= "3' untranslated region contg. the poly-A  
 signal of the nopaline-synthase gene of  
 Agrobacterium T-DNA"

FT complement (3032..3367)

FT /\*tag= f

FT /label= Barnase

FT /product= barnase

FT complement (3368..4876)

FT /\*tag= g

FT /label= PTA29

FT /note= "promoter region of tobacco TA29 gene"

FT complement (4822..4946)

FT /\*tag= h

FT /label= LB

FT /note= "T-DNA left border"

XX EP757102-A1.

XX PD 05-FEB-1997.

XX PF 04-AUG-1995; 95EP-0401844.

XX PR 04-AUG-1995; 95EP-0401844.

XX (PLBZ ) PLANT GENETIC SYSTEMS NV.

XX De Block M;

XX WPI; 1997-111050/11.

XX Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase  
 PT inhibitor - reduces the cultured cells response to stress and  
 PT reduces metabolism

XX Example 3; Page 13-16; 25pp; English.

XX Plasmid pTHW107 is a vector carrying T-DNA (AAT59531) comprising a  
 CC barnase coding sequence under control of the tobacco TA29 gene  
 CC stamen-specific promoter and a phosphinothricin acetyltransferase  
 CC coding sequence under control of an Arabidopsis Rubisco small  
 CC subunit gene promoter. Oilseed rape hypocotyl explants were  
 CC infected with Agrobacterium tumefaciens C58C1Rif carrying vector  
 CC pTHW107 and helper Ti plasmid pMP60. In some treatments, the  
 CC hypocotyls were treated with the poly-(ADP-ribose) polymerase  
 CC inhibitor niacinamide (250 mg/l) 4 days prior to infection.  
 CC Plants regenerated from niacinamide-treated transformed calli  
 CC had a low copy number and displayed less variation in the  
 CC expression profile of the transgenes.

XX Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match 100.0%; Score 25; DB 18; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.0078;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgcgatgagctaagctagc 25

Db 317 GGATCCCGCGATGAGCTAAGCTAGC 293

RESULT 9  
 AAH25422/c

XX ID AAH25422 standard; DNA; 4946 BP.

XX AC AAH25422;

XX DT 22-AUG-2001 (first entry)

XX Nucleotide sequence of plasmid pTHW107.

XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; ss.

XX Synthetic.

OS Streptomyces hygroscopicus.

OS Arabidopsis thaliana.

OS Bacillus amyloliquefaciens.

OS Nicotiana tabacum.

XX Key Location/Qualifiers

FT misc\_feature 1..25

FT /\*tag= a

FT /note= "right border repeat from TL-DNA from pTiB6S3"

FT misc\_feature 26..97

FT /\*tag= b

FT /note= "synthetic polylinker derived sequences"

```
FT 3'UTR complement (98...309)
FT /*tag= c
FT /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"
FT misc_feature 310...330
FT /*tag= d
FT CDS /note= "synthetic polylinker derived sequences"
FT 331...882
FT /*tag= e
FT /note= "Streptomyces hygrosopicus bar gene"
FT promoter complement (883...2608)
FT /*tag= f
FT /note= "ats1A ribulose-1,5-biphosphate carboxylase
FT misc_feature 2609...2658
FT /*tag= g
FT /note= "synthetic polylinker derived sequences"
FT 3'UTR complement (2659...2919)
FT /*tag= h
FT /note= "tagl fragment from 3' UTR of nopaline
FT 2920...3031
FT /*tag= i
FT /note= "3'UTR downstream of Bacillus amyloliquefaciens
FT CDS complement (3032...3367)
FT /*tag= j
FT /note= "Barnase coding region from Bacillus
FT promoter complement (3368...4877)
FT /*tag= k
FT /note= "anther-specific gene TA29 promoter from
FT misc_feature 4878...4921
FT /*tag= l
FT /note= "synthetic polylinker derived sequences"
FT misc_feature 4922...4946
FT /*tag= m
FT /note= "left border repeat from TL-DNA from pTiB6S3"
FT WO200141558-A1.
FT 14-JUN-2001.
FT 06-DEC-2000; 2000WO-EPI2872.
FT 08-DEC-1999; 99US-0457037.
FT (AVET ) AVENTIS CROPS SCIENCE NV.
FT De Both G, De Beuckeleer M;
FT WPI; 2001-381419/40.
FT Transgenic winter oilseed rape plants suited for producing hybrid seed
FT with improved qualities, comprises a male-sterility gene and fertility
FT restorer gene, integrated into the genome -
FT Example 1; Page 78-80; 98pp; English.
FT The specification describes a pair of transgenic winter oilseed rape
FT plants suited for producing hybrid seed. One of the plants has an
FT expression cassette comprising a male-sterility gene, and the other
FT plant has an expression cassette comprising a fertility restorer gene,
FT integrated into the genome. The fertility restorer gene is capable of
FT preventing the activity of the male-sterility gene. The plant pair is
FT useful for producing hybrid seed. Plants developed from the hybrid
FT seed have agronomic performance, genetic stability and adaptability to
FT different genetic backgrounds. The present sequence represents
FT a plasmid pTiW107. This plasmid comprises the barnase gene, which acts as
FT a male-sterility gene. The plasmid is used to create transgenic plants
FT of the invention.
XX
```

```
SQ Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;
Query Match 100.0%; Score 25; DB 22; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.0078;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggatcccccgatgagctaagctagc 25
| | | | | | | | | | | | | | | | | |
Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293
RESULT 10
AAF86439
ID AAF86439 standard; DNA; 5228 BP.
XX
AC AAF86439;
XX
XX 25-JUN-2001 (first entry)
XX Plasmid pTS172delta.
XX Male sterile plant; RNAase inhibitor; plasmid pTS172delta; ds.
XX Unidentified.
XX OS
XX WO200124616-A1.
XX PN
XX 12-APR-2001.
XX PD
XX 12-SEP-2000; 2000WO-JP06222.
XX PF
XX 30-SEP-1999; 99JP-0279307.
XX PR
XX (NISB ) JAPAN TOBACCO INC.
XX PA
XX Hamada K, Nakakido F;
XX PI
XX WPI; 2001-266212/27.
XX DR
XX Method for producing male sterile rice and maize by inserting RNAse
XX gene and RNAse inhibitor genes with promoters into the plant genome -
XX Disclosure; Page 14-17; 29pp; Japanese.
XX The present invention relates to a method for producing male sterile
XX plants. The method comprises inserting a promoter fragment upstream of an
XX RNAse gene and a second promoter, upstream of an RNAse inhibitor protein
XX gene and inserting it into the plant genome. The method is useful for
XX producing male sterile tobacco, lettuce and rapeseed plants, but
XX preferably rice and maize. The present sequence is a vector used in
XX the method of the present invention.
SQ Sequence 5228 BP; 1384 A; 1307 C; 1263 G; 1274 T; 0 other;
Query Match 100.0%; Score 25; DB 22; Length 5228;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggatcccccgatgagctaagctagc 25
| | | | | | | | | | | | | | | | | |
Db 4957 GGATCCCCCGATGAGCTAAGCTAGC 4981
RESULT 11
AAV23239/C
ID AAV23239 standard; DNA; 5349 BP.
XX
AC AAV23239;
XX
XX 17-JUL-1998 (first entry)
XX
```



DE T-DNA of pTTS24.  
XX Barstar; barnase inhibitor; fertility restoration;  
KW male-sterile line; plasmid pTTS24; T-DNA; ds.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature /tag= a  
FT /label= RB  
FT /note= "right boarder"  
FT complement (98..331)  
FT /tag= b  
FT /label= 3'-g7  
FT /note= "region containing 3' untranslated end of  
FT Agrobacterium T-DNA gene 7"  
FT CDS 332..883  
FT /tag= c  
FT /label= bar  
FT /note= "region coding for phosphinothricin acetyl  
FT transferase"  
FT complement (884..2258)  
FT /tag= d  
FT /label= P35S  
FT /note= "35S promoter of Cauliflower Mosaic Virus"  
FT 2281..3969  
FT /tag= e  
FT /label= PEl  
FT /note= "promoter of E1 gene of rice (W09213956)"  
FT 3970..4245  
FT /tag= f  
FT /product= improved\_barstar  
FT 4246..4577  
FT /tag= g  
FT /label= 3'\_chs  
FT /note= "region containing 3' untranslated end of  
FT chalcone synthase gene"  
FT 5325..5349  
FT /tag= h  
FT /note= "T-DNA left border"  
XX  
XX W09810081-A2.  
XX  
XX 12-MAR-1998.  
XX  
XX 01-SEP-1997; 97WO-EP04739.  
XX  
XX 03-SEP-1996; 96EP-0202446.  
XX  
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.  
XX Michiels F, Williams M;  
XX  
XX WPI; 1998-193630/17.  
XX  
XX DNA encoding an improved barstar protein - used to restore fertility  
XX in male-sterile plant lines  
XX  
XX Example 4; Pages 41-43; 54pp; English.  
XX  
XX The present sequence was used in the preparation of an improved  
CC Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which  
CC can be used to restore fertility to male-sterile lines.  
CC The DNA sequence encoding the improved barstar, leads to increased  
CC barstar production in tapetum cells, due to improved translation,  
CC and possibly protein stability.  
XX  
XX Sequence 5349 BP; 1339 A; 1233 C; 1290 G; 1487 T; 0 other;  
SQ  
Query Match 100.0%; Score 25; DB 19; Length 5349;  
Best Local Similarity 100.0%; Pred. No. 0.0079;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ggaatccccccgatgagctagctagc 25  
Db 317 GGATCCCCCGATGAGCTAGCTAGC 293  
RESULT 12  
AAT39339  
ID AAT39339 standard; DNA; 5864 BP.  
XX  
XX AAT39339;  
XX  
XX 22-JAN-1997 (first entry)  
XX  
XX Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.  
DE Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;  
KW transgenic plant; oilseed rape; canole; Brassica napus; ds.  
KW  
XX Synthetic.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature /tag= a  
FT /label= RB  
FT /note= "right border of Agrobacterium T-DNA"  
FT complement (98..330)  
FT /tag= b  
FT /label= 3'-g7  
FT /note= "region containing polyA signal of gene 7  
FT of Agrobacterium T-DNA"  
FT complement (331..882)  
FT /tag= c  
FT /label= bar  
FT /note= "region coding for phosphinothricin  
FT acetyltransferase"  
FT complement (883..2608)  
FT /tag= d  
FT /label= Pssu  
FT /note= "promoter of Arabidopsis Rubisco small  
FT subunit gene"  
FT complement (2659..3031)  
FT /tag= e  
FT /label= 3'nos  
FT /note= "region containing polyA signal of nopaline  
FT synthase gene of Agrobacterium T-DNA"  
FT complement (3032..3367)  
FT /tag= f  
FT /label= Barnase  
FT /note= "Bacillus amyloliquefaciens barnase coding  
FT region"  
FT complement (3368..4877)  
FT /tag= g  
FT /label= PTA29  
FT /note= "promoter of stamen-specific TA29 gene of  
FT Nicotiana tabacum"  
FT 4924..5216  
FT /tag= h  
FT /label= Pnos  
FT /note= "promoter of nopaline synthase gene of  
FT Agrobacterium T-DNA"  
FT 5217..5489  
FT /tag= i  
FT /label= Barstar  
FT /note= "region coding for barstar of Bacillus  
FT amyloliquefaciens"  
FT 5490..5765  
FT /tag= j  
FT /label= 3'-g7  
FT /note= "region containing polyA signal of gene 7  
FT of Agrobacterium T-DNA"  
FT complement (5840..5864)  
FT misc\_feature

```

FT      /*tag= k
FT      /label= LB
FT      /note= "left border of Agrobacterium T-DNA"
XX      WO9626283-A1.
XX      29-AUG-1996.
XX      21-FEB-1996; 96WO-EP00722.
XX      21-FEB-1995; 95EP-0400364.
XX      (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX      Botterman J, Cornelissen M, Michiels F;
XX      WPI; 1996-402373/40.
XX      Prodn. of male sterile plants by transforming with a chimaeric
XX      construct - comprising a male sterility DNA e.g. barnase and a
XX      co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX      for generating hybrid cultivars
XX      Example 3; Page 33-3743-47; 56pp; English.
XX      Plasmid pTCO113 (AAT39339) is a T-DNA vector containing a bar gene
XX      under control of the PSU promoter, a barnase gene under control
XX      of the stamen-specific PRA29 promoter, and a barstar (co-regulatory)
XX      gene under control of the Pnos promoter. 87% Of oilseed rape
XX      plants regenerated after Agrobacterium-mediated transformation
XX      using pTCO113 were male sterile. Barnase expression disturbed the
XX      function of stamen cells leading to male sterility. Constitutive
XX      expression of barstar counteracted any low level expression of
XX      barnase in non-stamen tissue.
XX      Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;
XX      Query Match 100.0%; Score 25; DB 17; Length 5864;
XX      Best Local Similarity 100.0%; Pred. No. 0.0079;
XX      Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX      Qy 1 ggatcccccgatgagctaaagctagc 25
XX      |||||||
XX      Db 5546 ggatcccccgatgagctaaagctagc 5570
XX      RESULT 13
XX      AAT39339/c
XX      AC AAT39339 standard; DNA; 5864 BP.
XX      AC AAT39339;
XX      DT 22-JAN-1997 (first entry)
XX      DE Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.
XX      KW Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;
XX      KW transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX      OS Synthetic.
XX      FH Key Location/Qualifiers
XX      misc_feature complement (1..25)
XX      /tag= a
XX      /label= RB
XX      /note= "right border of Agrobacterium T-DNA"
XX      polyA_signal complement (98..330)
XX      /tag= b
XX      /label= 3'g7
XX      /note= "region containing polyA signal of gene 7
XX      of Agrobacterium T-DNA"
XX      CDS complement (331..882)

```

```

FT      /*tag= c
FT      /label= bar
FT      /note= "region coding for phosphinothricin
FT      acetyltransferase"
XX      complement (883..2608)
XX      /tag= d
XX      /label= Psu
XX      /note= "promoter of Arabidopsis Rubisco small
XX      subunit gene"
XX      complement (2659..3031)
XX      /tag= e
XX      /label= 3'nos
XX      /note= "region containing polyA signal of nopaline
XX      synthase gene of Agrobacterium T-DNA"
XX      CDS complement (3032..3367)
XX      /tag= f
XX      /label= Barnase
XX      /note= "Bacillus amyloliquefaciens barnase coding
XX      region"
XX      complement (3368..4877)
XX      /tag= g
XX      /label= PRA29
XX      /note= "promoter of stamen-specific TA29 gene of
XX      Nicotiana tabacum"
XX      4924..5216
XX      /tag= h
XX      /label= Pnos
XX      /note= "promoter of nopaline synthase gene of
XX      Agrobacterium T-DNA"
XX      CDS 5217..5489
XX      /tag= i
XX      /label= Barstar
XX      /note= "region coding for barstar of Bacillus
XX      amyloliquefaciens"
XX      5490..5765
XX      /tag= j
XX      /label= 3'g7
XX      /note= "region containing polyA signal of gene 7
XX      of Agrobacterium T-DNA"
XX      misc_feature complement (5840..5864)
XX      /tag= k
XX      /label= LB
XX      /note= "left border of Agrobacterium T-DNA"
XX      WO9626283-A1.
XX      PN
XX      PD 29-AUG-1996.
XX      PF 21-FEB-1996; 96WO-EP00722.
XX      XX 21-FEB-1995; 95EP-0400364.
XX      (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX      PA
XX      PI Botterman J, Cornelissen M, Michiels F;
XX      WPI; 1996-402373/40.
XX      DR Prodn. of male sterile plants by transforming with a chimaeric
XX      construct - comprising a male sterility DNA e.g. barnase and a
XX      co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX      for generating hybrid cultivars
XX      Example 3; Page 33-3743-47; 56pp; English.
XX      Plasmid pTCO113 (AAT39339) is a T-DNA vector containing a bar gene
XX      under control of the PSU promoter, a barnase gene under control
XX      of the stamen-specific PRA29 promoter, and a barstar (co-regulatory)
XX      gene under control of the Pnos promoter. 87% Of oilseed rape
XX      plants regenerated after Agrobacterium-mediated transformation
XX      using pTCO113 were male sterile. Barnase expression disturbed the
XX      function of stamen cells leading to male sterility. Constitutive
XX      expression of barstar counteracted any low level expression of

```

```
CC barnase in non-stamen tissue.
XX
SQ Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;

Query Match      100.0%; Score 25; DB 17; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggatcccccgcgactgaactagc 25
    |||||
Db 317 GGATCCCCCGAGTAACTAGC 293

RESULT 14
AAD06990
ID AAD06990 standard; DNA; 5865 BP.
XX
AC AAD06990;
XX
DT 06-AUG-2001 (first entry)
XX
DE Chimeric T-DNA of plasmid pTCO113.
XX
KW T-DNA; plasmid pTCO113; transgenic Brassica plant; transformation event;
KW male-sterility gene; chimeric; tobacco; ds.
XX
OS Chimeric - Streptomyces hygroscopicus.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Bacillus amyloliquefaciens.
OS Chimeric - Nicotiana tabacum.
OS Chimeric - Agrobacterium tumefaciens.
OS Chimeric - Unidentified.
XX
PH Key Location/Qualifiers
FT misc_feature 1..25
FT /note= "Right border repeat from the TL-DNA from
FT pTiB6S3"
FT 26..53
FT misc_feature
FT /tag= b
FT /note= "Synthetic polylinker derived sequence"
FT 54..90
FT misc_feature
FT /tag= c
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 98..309
FT 3'UTR
FT /tag= d
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT 310..331
FT misc_feature
FT /tag= e
FT /note= "Synthetic polylinker derived sequence"
FT 332..883
FT CDS
FT /tag= f
FT /product= "Protein encoded by bialaphos resistance
FT gene (bar) of Streptomyces hygroscopicus"
FT 884..2609
FT promoter
FT /tag= g
FT /note= "Promoter from the atSLA ribulose-1,5-bisphosphate
FT carboxylase small subunit gene from Arabidopsis thaliana"
FT 2610..3659
FT misc_feature
FT /tag= h
FT /note= "Synthetic polylinker derived sequence"
FT 2660..2920
FT misc_feature
FT /tag= i
FT /note= "Tag1 fragment from the 3' untranslated end of the
FT nopaline synthase gene (3'nos) from the T-DNA of pTi37
FT and containing plant polyadenylation signals"
FT 2921..2936
FT misc_feature
FT /tag= j
FT /note= "Synthetic polylinker derived sequence"
FT 2937..3032
FT 3'UTR
FT
```

```
FT /tag= k
FT /note= "The 3' untranslated region downstream from the
FT barnase coding sequence of Bacillus amyloliquefaciens"
FT 3033..3368
FT CDS
FT /tag= l
FT /product= "Protein encoded by barnase gene from
FT Bacillus amyloliquefaciens"
FT 3369..4878
FT promoter
FT /tag= m
FT /note= "Promoter region of the anther-specific gene
FT TA29 from Nicotiana tabacum"
FT 4879..4924
FT misc_feature
FT /tag= n
FT /note= "Synthetic polylinker derived sequence"
FT 4925..5215
FT promoter
FT /tag= o
FT /note= "Promoter of the nopaline synthase gene from the
FT T-DNA of pTi37 of Agrobacterium tumefaciens"
FT 5216..5217
FT misc_feature
FT /tag= p
FT /note= "Synthetic polylinker derived sequence"
FT 5218..5490
FT CDS
FT /tag= q
FT /product= "Protein encoded by barstar gene of
FT Bacillus amyloliquefaciens"
FT 5491..5530
FT misc_feature
FT /tag= r
FT /note= "Sequence from the 3' untranslated end of the
FT barstar gene from Bacillus amyloliquefaciens"
FT 5531..5554
FT misc_feature
FT /tag= s
FT /note= "Synthetic polylinker derived sequence"
FT 5555..5766
FT 3'UTR
FT /tag= t
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT 5767..5773
FT misc_feature
FT /tag= u
FT /note= "Synthetic polylinker derived sequence"
FT 5774..5810
FT misc_feature
FT /tag= v
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 5811..5840
FT misc_feature
FT /tag= w
FT /note= "Synthetic polylinker derived sequence"
FT 5841..5865
FT misc_feature
FT /tag= x
FT /note= "Left border repeat from the TL-DNA from
FT pTiB6S3"
FT
XX WO200131042-A2.
XX
PD 03-MAY-2001.
XX
PF 26-OCT-2000; 2000WO-EP10680.
XX
XX 29-OCT-1999; 99US-0430497.
PR
XX (AVET ) AVENTIS CROPS SCIENCE NV.
PA
XX Weston B, De Beuckeleer M;
PI
XX WPI; 2001-300517/31.
DR
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by
PT harboring specific transformation events, particularly by presence of
PT male-sterility gene, at specific location in its genome.
XX
PS Claim 1; Page 47-49; 53pp; English.
CC The present invention relates to a transgenic Brassica plant or its
CC seed, cells or tissues, characterised by harbouring a specific
CC
```

```

CC transformation event, particularly by the presence of a male-sterility
CC gene, at a specific location in the Brassica genome. Transgenic
CC Brassica plant is useful for producing a hybrid seed by crossing the
CC transgenic plant with a male-fertile Brassica plant and harvesting the
CC hybrid seed from the transgenic Brassica plant.
CC The present sequence is chimeric r-DNA of plasmid pTColl3. This sequence
CC comprises right border repeat, left border repeat and 3' untranslated
CC region (UTR) from TL-DNA of pTiB6S3, synthetic polylinker sequences,
CC coding regions of bialaphos resistance gene (bar) from
CC Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens
CC and barstar gene from Bacillus amyloliquefaciens and promoters of atSIA
CC ribulose-1,5-biphosphate carboxylase small subunit gene from
CC Arabidopsis thaliana, the anther-specific gene TA29 from
CC Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTiT37
CC of Agrobacterium tumefaciens.
XX
SQ Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;

Query Match          100.0%; Score 25; DB 22; Length 5865;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgcagctgaagctagc 25
   |||||||
Db 5547 ggatcccccgcagctgaagctagc 5571

RESULT 15
AAD06990/c
ID AAD06990 standard; DNA; 5865 BP.
AC AAD06990;
XX
DT 06-AUG-2001 (first entry)
XX
DE Chimeric T-DNA of plasmid pTColl3.
XX
KW T-DNA; plasmid pTColl3; transgenic Brassica plant; transformation event;
KW male-sterility gene; chimeric; tobacco; ds.
XX
OS Chimeric - Streptomyces hygroscopicus.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Bacillus amyloliquefaciens.
OS Chimeric - Nicotiana tabacum.
OS Chimeric - Agrobacterium tumefaciens.
XX
XX Chimeric - Unidentified.
FH
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "Right border repeat from the TL-DNA from
FT pTiB6S3"
FT 26..53
FT misc_feature
FT /tag= b
FT /note= "Synthetic polylinker derived sequence"
FT 54..90
FT misc_feature
FT /tag= c
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 98..309
FT 3'UTR
FT /tag= d
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT 310..331
FT misc_feature
FT /tag= e
FT /note= "Synthetic polylinker derived sequence"
FT 332..883
FT CDS
FT /tag= f
FT /product= "Protein encoded by bialaphos resistance
FT gene (bar) of Streptomyces hygroscopicus"
FT 884..2609
FT /tag= g

```

```

FT /note= "Promoter from the atSIA ribulose-1,5-biphosphate
FT carboxylase small subunit gene from Arabidopsis thaliana"
FT 2610..2659
FT /tag= h
FT /note= "Synthetic polylinker derived sequence"
FT 2660..2920
FT /tag= i
FT /note= "TaqI fragment from the 3' untranslated end of the
FT nopaline synthase gene (3'nos) from the T-DNA of pTiT37
FT and containing plant polyadenylation signals"
FT 2921..2936
FT /tag= j
FT /note= "Synthetic polylinker derived sequence"
FT 2937..3032
FT /tag= k
FT /note= "The 3' untranslated region downstream from the
FT barnase coding sequence of Bacillus amyloliquefaciens"
FT 3033..3368
FT /tag= l
FT /product= "Protein encoded by barnase gene from
FT Bacillus amyloliquefaciens"
FT 3369..4878
FT /tag= m
FT /note= "Promoter region of the anther-specific gene
FT TA29 from Nicotiana tabacum"
FT 4879..4924
FT /tag= n
FT /note= "Synthetic polylinker derived sequence"
FT 4925..5215
FT /tag= o
FT /note= "Promoter of the nopaline synthase gene from the
FT T-DNA of pTiT37 of Agrobacterium tumefaciens"
FT 5216..5217
FT /tag= p
FT /note= "Synthetic polylinker derived sequence"
FT 5218..5490
FT /tag= q
FT /product= "Protein encoded by barstar gene of
FT Bacillus amyloliquefaciens"
FT 5491..5530
FT /tag= r
FT /note= "Sequence from the 3' untranslated end of the
FT barstar gene from Bacillus amyloliquefaciens"
FT 5531..5554
FT /tag= s
FT /note= "Synthetic polylinker derived sequence"
FT 5555..5766
FT /tag= t
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT 5767..5773
FT /tag= u
FT /note= "Synthetic polylinker derived sequence"
FT 5774..5810
FT /tag= v
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 5811..5840
FT /tag= w
FT /note= "Synthetic polylinker derived sequence"
FT 5841..5865
FT /tag= x
FT /note= "Left border repeat from the TL-DNA from
FT pTiB6S3"
FT
XX WO200131042-A2.
XX 03-MAY-2001.
XX 26-OCT-2000; 2000WO-EP10680.
XX 29-OCT-1999; 99US-0430497.
XX

```

PA (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 PI Weston B, De Beuckeleer M;  
 XX  
 DR WPI; 2001-300517/31.  
 XX  
 PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX  
 XX  
 PS Claim 1; Page 47-49; 53pp; English.  
 XX  
 CC The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is chimeric T-DNA of plasmid pTCO113. This sequence  
 CC comprises right border repeat, left border repeat and 3' untranslated  
 CC region (UTR) from TL-DNA of pTiB6S3, synthetic polylinker sequences,  
 CC coding regions of blalaphos resistance gene (bar) from  
 CC Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens  
 CC and barstar gene from Bacillus amyloliquefaciens and promoters of ats1A  
 CC ribulose-1,5-bisphosphate carboxylase small subunit gene from  
 CC Arabidopsis thaliana, the anther-specific gene TA29 from  
 CC Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTiT37  
 CC of Agrobacterium tumefaciens.  
 XX  
 SQ Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 5865;  
 Best Local Similarity 100.0%; Pred. No. 0.0079;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ggatcccccgatgagctaaagctagc 25  
 |||||  
 Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

Search completed: February 15, 2002, 19:01:18  
 Job time: 20722 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:07:28 ; Search time 353.79 Seconds  
(without alignments)  
16.004 Million cell updates/sec

Title: US-09-698-903B-5

Perfect score: 25  
Sequence: 1 ggatccccgatgagctaagctagc 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1303	US-08-894-440-2	Sequence 2, Appli
c 2	25	100.0	4946	US-08-817-188-1	Sequence 1, Appli
c 3	25	100.0	5560	US-08-817-188-5	Sequence 5, Appli
4	25	100.0	5864	US-08-894-440-4	Sequence 4, Appli
c 5	25	100.0	5864	US-08-894-440-4	Sequence 4, Appli
6	25	100.0	6548	US-08-894-440-1	Sequence 1, Appli
7	25	100.0	6548	US-08-817-188-2	Sequence 2, Appli
8	24	96.0	3200	US-08-453-104-23	Sequence 23, Appli
9	24	96.0	3200	US-08-694-824-23	Sequence 23, Appli
10	24	96.0	3201	US-08-453-104-22	Sequence 22, Appli
11	24	96.0	3201	US-08-694-824-22	Sequence 22, Appli
12	24	96.0	7566	US-08-232-016-22	Sequence 23, Appli
13	24	96.0	7639	US-08-232-016-22	Sequence 23, Appli
c 14	24	96.0	7811	US-08-549-680A-5	Sequence 2, Appli
c 15	17	68.0	29	US-08-232-016-11	Sequence 5, Appli
16	17	68.0	3153	US-09-080-625-3	Sequence 3, Appli
17	17	68.0	3336	US-09-080-625-2	Sequence 2, Appli
18	17	68.0	3694	US-09-080-625-5	Sequence 5, Appli
19	17	68.0	3877	US-09-080-625-4	Sequence 4, Appli
c 20	17	68.0	24595	5428147-1	Patent No. 5428147
c 21	16.6	66.4	6201	US-08-790-912-1	Sequence 1, Appli
22	16	64.0	54	US-08-503-226B-32	Sequence 32, Appli
23	16	64.0	54	US-08-721-458B-32	Sequence 32, Appli
c 24	16	64.0	3065	US-08-209-521-28	Sequence 2, Appli
25	15.4	61.6	1186	US-08-064-121-2	Sequence 2, Appli
26	15.4	61.6	1186	US-08-478-015-2	Sequence 2, Appli
27	15.4	61.6	1186	US-08-475-975-2	Sequence 2, Appli

28	15.4	61.6	1186	3	US-09-084-889-2	Sequence 2, Appl
29	15.4	61.6	1859	3	US-08-691-563C-46	Sequence 46, Appl
c 30	15.4	61.6	2143	2	US-08-656-177A-1	Sequence 1, Appl
c 31	15.4	61.6	2143	3	US-09-256-797-1	Sequence 1, Appl
c 32	15.4	61.6	10281	2	US-08-816-155B-1	Sequence 1, Appl
c 33	15.4	61.6	10281	3	US-09-079-587-1	Sequence 1, Appl
c 34	15.2	60.8	43795	3	US-08-742-185-101	Sequence 101, App
35	15	60.0	1642	1	US-07-996-772A-1	Sequence 1, Appl
36	15	60.0	1642	1	US-08-446-822-1	Sequence 1, Appl
37	15	60.0	1642	5	PCT-US93-12586-1	Sequence 1, Appl
c 38	15	60.0	2352	2	US-08-922-837-1	Sequence 1, Appl
c 39	15	60.0	2352	4	US-09-351-550-1	Sequence 1, Appl
c 40	15	60.0	7431	4	US-09-306-398-2	Sequence 2, Appl
c 41	14.8	59.2	1377	1	US-08-112-817C-1	Sequence 1, Appl
42	14.8	59.2	1710	4	US-09-068-960-14	Sequence 14, Appl
43	14.6	58.4	376	3	US-08-985-950-3	Sequence 3, Appl
c 44	14.6	58.4	699	4	US-08-998-416-591	Sequence 591, App
c 45	14.4	57.6	60	4	US-09-171-945-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1  
US-08-894-440-2  
; Sequence 2, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894.440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1303  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI  
; OTHER INFORMATION: fragment of pT588  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(35)  
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36)..(694)  
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain  
; OTHER INFORMATION: CM1841 (p35S)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (695)..(967)  
; OTHER INFORMATION: region coding for barstar of Bacillus  
; OTHER INFORMATION: amyloliquefaciens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (968)..(1287)  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1288)..(1303)  
; OTHER INFORMATION: polylinker of pGEM2  
; OTHER INFORMATION: polylinker of pGEM2  
US-08-894-440-2

Query Match 100.0%; Score 25; DB 3; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggatccccgatgagctaagctagc 25

Db 1024 ggatccccgatgagctagctagc 1048  
 |||

## RESULT 2

```

US-08-817-188-1/c
; Sequence 1, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4946
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW107
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: T-DNA right border (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((97)..(330))
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl
; OTHER INFORMATION: transferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter region of Rubisco small subunit gene of
; OTHER INFORMATION: Arabidopsis thaliana (PSSU)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2658)..(3031))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4876))
; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((4922)..(4946))
; OTHER INFORMATION: LB: T-DNA left border
Us-08-817-188-1

```

Query Match 100.0%; Score 25; DB 3; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.0014;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatccccgatgagctagctagc 25  
 |||  
 Db 317 GGATCCCGGATGAGCTAAGCTAGC 293

## RESULT 3

```

US-08-817-188-5/c
; Sequence 5, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CaMV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-LS1 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5067)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CaMV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: PTB6S3
; FEATURE:

```



```
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; US-08-817-188-5

Query Match      100.0%; Score 25; DB 3; Length 5560;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatccccgatgagctaaagctagc 25
   |||||
Db 304 GGATCCCCGATGAGCTAAGCTAGC 280

RESULT 4
US-08-894-440-4
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTColl3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
```

```
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
; US-08-894-440-4

Query Match      100.0%; Score 25; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatccccgatgagctaaagctagc 25
   |||||
Db 5546 ggaatccccgatgagctaaagctagc 5570

RESULT 5
US-08-894-440-4/c
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTColl3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
```

```

; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Phos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4

Query Match      100.0%; Score 25; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
Db 317 GGATCCCCGATGAGCTAAGCTAGC 293

RESULT 6
; Sequence 1, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTS174
; NAME/KEY: misc_feature
; LOCATION: (1)..(2003)
; OTHER INFORMATION: pUC19 derived vector sequences (vector)
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2283))
; OTHER INFORMATION: 3' nos: region containing polyadenylation signal
; OTHER INFORMATION: of nopaline synthase gene of Agrobacterium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2284)..(2624))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: promoter of the stamen-specific El gene of rice
; OTHER INFORMATION: (PE1)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5710)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus (P35S)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5711)..(6262)
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: region containing polyadenylation signal fo gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
US-08-894-440-1

Query Match      100.0%; Score 25; DB 3; Length 6548;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
Db 6277 ggatcccccgatgagctaaagctagc 6301

RESULT 7
US-08-817-188-2
; Sequence 2, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTS172
; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2288))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2289)..(2624))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:

```

; NAME/KEY: misc\_feature  
; LOCATION: Complement((2625)..(4313))  
; OTHER INFORMATION: P1: promoter region of E1 gene of rice  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4336)..(5170)  
; OTHER INFORMATION: P35S: 35S promoter region of Cauliflower Mosaic  
; OTHER INFORMATION: Virus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((5711)..(6262))  
; OTHER INFORMATION: bar: region coding for phosphinothricin  
; OTHER INFORMATION: acetyltransferase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (6263)..(6496)  
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
; OTHER INFORMATION: T-DNA  
US-08-817-188-2

Query Match 100.0%; Score 25; DB 3; Length 6548;  
Best Local Similarity 100.0%; Pred. No. 0.0014; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0;

Qy 1 ggatcccccgatgagctaagctagc 25  
|||||  
Db 6277 ggatcccccgatgagctaagctagc 6301

RESULT 8  
US-08-453-104-23  
; Sequence 23, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453.104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-6620  
; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-23

Query Match 96.0%; Score 24; DB 1; Length 3200;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gatcccccgatgagctaagctagc 25  
|||||  
Db 2879 GATCCCGGATGAGCTAAGCTAGC 2902

RESULT 9  
US-08-694-824-23  
; Sequence 23, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694.824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature

LOCATION: 2078..2082  
OTHER INFORMATION: /note= "Nucleotides 2078-2082  
wherein N is not known."  
US-08-694-824-23

Query Match 96.0%; Score 24; DB 2; Length 3200;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatccccgatgagctaaagctagc 25  
|||||  
Db 2879 GATCCCCGATGAGCTAAGCTAGC 2902

RESULT 10  
US-08-453-104-22  
; Sequence 22, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
wherein N is not known."  
US-08-453-104-22

Query Match 96.0%; Score 24; DB 1; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 0.004;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 gatccccgatgagctaaagctagc 25  
|||||  
Db 2952 GATCCCCGATGAGCTAAGCTAGC 2975

RESULT 11  
US-08-694-824-22  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
wherein N is not known."  
US-08-694-824-22

Query Match 96.0%; Score 24; DB 2; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatccccgatgagctaaagctagc 25  
|||||  
Db 2952 GATCCCCGATGAGCTAAGCTAGC 2975

RESULT 12

```

US-08-232-016-23
; Sequence 23, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91402920.2
; FILING DATE: 30-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pPS0212
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated modified bt2 (cryIab) gene, also designated as the
; OTHER INFORMATION: cryIab6 gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1793..2026
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2396..2921
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2922..3581
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: 3582..4407
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobact
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5600..6457
; OTHER INFORMATION: /note= "Sequence complementary to
; Patent No. 5952547
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7071..7566
; OTHER INFORMATION: /note= "TR1' and TR2 promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA (with modified leader wi
; OTHER INFORMATION: respect to sequence of pJD884 of SEQ ID NO. 22."
US-08-232-016-23

Query Match          96.0%; Score 24; DB 2; Length 7566;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatcccccgatgagctaaagctagc 25
|||||
Db 1808 GATCCCCCGATGAGCTAAGCTAGC 1831

RESULT 13
US-08-232-016-22
; Sequence 22, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; TITLE OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pPS0212
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated modified bt2 (cryIab) gene, also designated as the
; OTHER INFORMATION: cryIab6 gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1793..2026
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2396..2921
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2922..3581
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:

```

```

; LENGTH: 7639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pJD884
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1869
;   OTHER INFORMATION: /note= "Coding region of a
;   OTHER INFORMATION: truncated bt2 (cryIab) gene, also designated as the bt884 gene"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1877..2110
;   OTHER INFORMATION: /note= "3' regulatory sequence
;   OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium
;   OTHER INFORMATION: T-DNA gene 7."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2480..3005
;   OTHER INFORMATION: /note= "35S promoter sequence
;   OTHER INFORMATION: derived from Cauliflower mosaic virus."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3006..3665
;   OTHER INFORMATION: /note= "Coding sequence of
;   OTHER INFORMATION: chloramphenicol acetyl transferase gene."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3666..4491
;   OTHER INFORMATION: /note= "3' regulatory sequence
;   OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium
;   OTHER INFORMATION: T-DNA octopine synthase gene."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 5684..6541
;   OTHER INFORMATION: /note= "Sequence complementary to
;   OTHER INFORMATION: Patent No. 5952547
;   OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 7155..7639
;   OTHER INFORMATION: /note= "TR1' and TR2' promoter
;   OTHER INFORMATION: derived from Agrobacterium T-DNA."
;
US-08-232-016-22

```

```

Query Match          96.0%; Score 24; DB 2; Length 7639;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 gatccccgatgagctagctagc 25
   |||||
Db 1892 GATCCCCGATGAGCTAGCTAGC 1915

```

```

RESULT 14
US-08-549-680A-5/c
; Sequence 5, Application US/08549680A
; Patent No. 5962768
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, MARCUS
; APPLICANT: REYNALTS, ARLETTE
; APPLICANT: GOSSELE, VERONIQUE
; APPLICANT: VAN AARSSEN, ROEL
; TITLE OF INVENTION: MARKER GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia

```

```

; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,680A
; FILING DATE: 16 JANUARY 1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-0111P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7811 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (synthetic)
; FEATURE:
;   NAME/KEY: misc_recomb
;   LOCATION: 1..7811
;   OTHER INFORMATION: /label= vector pTRVA3
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 194..218
;   OTHER INFORMATION: /note= "T-DNA right border"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 484..684
;   OTHER INFORMATION: /note= "the 3' end formation and
;   OTHER INFORMATION: polyadenylation region of T-DNA gene 7"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: complement (729..1340)
;   OTHER INFORMATION: /note= "the aac(6') coding
;   OTHER INFORMATION: sequence"
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: 1341..1756
;   OTHER INFORMATION: /label= 35S promoter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3001..3023
;   OTHER INFORMATION: /note= "T-DNA left border
;   OTHER INFORMATION: sequences"
;
US-08-549-680A-5

```

```

Query Match          96.0%; Score 24; DB 2; Length 7811;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 gatccccgatgagctagctagc 25
   |||||
Db 702 GATCCCCGATGAGCTAGCTAGC 679

```

```

RESULT 15
US-08-232-016-11/c
; Sequence 11, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke

```

APPLICANT: DOCKX, Jan  
APPLICANT: VAN AARSEN, Roel  
TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN  
TITLE OF INVENTION: PLANT CELLS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,016  
FILING DATE: 03-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91402920.2  
FILING DATE: 30-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 92400820.4  
FILING DATE: 25-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K  
REGISTRATION NUMBER: P39,300  
REFERENCE/DOCKET NUMBER: 010830-049  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: oligonucleotide, designated as PS59  
US-08-232-016-11

Query Match 68.0%; Score 17; DB 2; Length 29;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gatccccgatgagcta 18  
|||||  
Db 17 GATCCCCGATGAGCTA 1

Search completed: February 15, 2002, 19:07:32  
Job time: 14626 sec

**THIS PAGE BLANK (USPIC)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:03:47 ; Search time 9904.61 Seconds  
(without alignments)  
27.123 Million cell updates/sec

Title: US-09-698-903B-5  
Perfect score: 25  
Sequence: 1 ggatcccccagatgagctaagctagc 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estfun:  
2: em\_esthum:  
3: em\_estin:  
4: em\_estom:  
5: em\_estpl:  
6: em\_estba:  
7: em\_estro:  
8: em\_estov:  
9: em\_htc:  
10: gb\_estl:  
11: gb\_est2:  
12: gb\_htc:  
13: gb\_gss:  
14: em\_gss\_fun:  
15: em\_gss\_hum:  
16: em\_gss\_inv:  
17: em\_gss\_pln:  
18: em\_gss\_pro:  
19: em\_gss\_rod:  
20: em\_gss\_vrt:  
21: em\_gss\_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.2	72.8	498	13	AQ791259 HS_5498_A
C 2	18.2	72.8	562	13	AZ028032 RPCI-23-3
C 3	17.6	70.4	303	10	AW862033 RC3-CT034
C 4	17.6	70.4	312	10	BB453584
C 5	17.6	70.4	454	11	N24161
6	17.6	70.4	497	11	BF496895
C 7	17.6	70.4	510	11	BI038578
8	17.6	70.4	587	11	BF877912
9	17.6	70.4	605	10	BE639934
10	17.6	70.4	611	13	AQ834638 HS_5442_A
C 11	17.6	70.4	924	13	AL233215 Tetraodon
C 12	17.2	68.8	295	10	BE120999 UI-R-CA0-

13	17.2	68.8	326	10	BB499943
C 14	17.2	68.8	336	11	BF410724 UI-R-CA0-
15	17.2	68.8	368	13	AQ867402 nbeb0031F
16	17.2	68.8	387	10	AW501051 UI-HF-BP0
C 17	17.2	68.8	398	10	AL183978 qd59901.x
C 18	17.2	68.8	407	10	AA429095 zV49d03.r
C 19	17.2	68.8	410	10	AW581963 MR4-ST012
C 20	17.2	68.8	462	11	BI134877 UI-M-BH3-
C 21	17.2	68.8	501	11	BG073283 H3120G12-
C 22	17.2	68.8	549	13	TA311E12Q
C 23	17.2	68.8	554	11	BI133671 UI-M-BH3-
C 24	17.2	68.8	554	13	AQ509694 nbxb0096F
C 25	17.2	68.8	568	10	BE048724 hr52h07.x
C 26	17.2	68.8	574	10	BE026535 db27g06.x
C 27	17.2	68.8	583	11	BI134114 UI-M-BH3-
C 28	17.2	68.8	585	10	BE326762 hr64e06.x
C 29	17.2	68.8	622	10	AW548994 L0045H03-
30	17.2	68.8	627	10	BE535820 601060510
C 31	17.2	68.8	630	11	BG470582 602511491
C 32	17.2	68.8	707	11	BG334339 602461858
C 33	17.2	68.8	729	10	BE394825 601312066
C 34	17.2	68.8	755	10	BE543444 601070751
C 35	17.2	68.8	768	11	EG678254 602624567
36	17.2	68.8	781	10	AI344042 tc01g05.x
37	17.2	68.8	812	11	BG035724 602325804
38	17.2	68.8	825	11	BG166218 602340976
39	17.2	68.8	894	11	BF301141 602029166
40	17.2	68.8	916	11	BF797773 602257530
41	17.2	68.8	992	13	CNS03330
C 42	17.2	68.8	1498	11	BG476758 602524680
C 43	17.2	68.8	1649	12	AK016762 mus muscu
C 44	17	68.0	251	10	BE419550 WWS014.E1
45	17	68.0	400	10	BE119925 UI-R-CA0-

ALIGNMENTS

RESULT 1  
AQ791259 498 bp DNA GSS 03-AUG-1999  
LOCUS HS\_5498\_A2\_B10\_T7A RPCI-11 Human Male BAC Library Homo sapiens  
DEFINITION genomic clone Plates=1074 Col=20 Row=C, DNA sequence.  
ACCESSION AQ791259  
VERSION AQ791259.1 GI:5698806  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 498)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.hscsc.washington.edu  
Plate: 1074 row: C column: 20  
Seq primer: T7

```

Class: BAC ends
High quality sequence stop: 498.
Location/Qualifiers
1. .498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate-1074 Col=20 Row=C"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      166 a   90 c   85 g   151 t       6 others
ORIGIN

Query Match      72.8%; Score 18.2; DB 13; Length 498;
Best Local Similarity 87.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atccccgatgagctaaagctagc 25
|||||  |||||  |||||  |||||  |||||
Db 414 ATCCCCCGATGTCCTAGATAGC 436

RESULT 2
AZ028032/c
LOCUS      562 bp      DNA      GSS      25-FEB-2000
DEFINITION RPCI-23-373M15_TV RPCI-23 Mus musculus genomic clone RPCI-23-373M15
          DNA sequence.
ACCESSION  AZ028032
VERSION     AZ028032.1 GI:7103416
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 562)
Zhao,S., Nierman,W., Feildblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-373M15.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact pter de Jong
(pter@edj.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 373 row: M column: 15
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. .562
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-373M15"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      122 a   122 c   163 g   153 t       2 others
ORIGIN

Query Match      72.8%; Score 18.2; DB 13; Length 562;
Best Local Similarity 87.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atccccgatgagctaaagctagc 25
|||||  |||||  |||||  |||||  |||||
Db 247 ATCCACCATGAGCTTAGCTAGC 225

RESULT 3
AW862033/c
LOCUS      303 bp      mRNA      EST      19-MAY-2000
DEFINITION RC3-CT0347-110300-014-a10 CT0347 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW862033
VERSION     AW862033.1 GI:7957726
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 303)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Balg,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC3-CT0347-110
300-014-a10&t3=2000-03-11&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 303.
Location/Qualifiers
1. .303
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0347"
/dev_stage="Adult"
/notes="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      104 a   60 c   62 g   77 t
ORIGIN

Query Match      70.4%; Score 17.6; DB 10; Length 303;

```



(Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS974) was kindly provided by Dr. Anthony P. Albino.  
 BASE COUNT 173 a 81 c 61 g 139 t  
 ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 454;  
 Best Local Similarity 83.3%; Pred. No. 5.3e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctag 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 448 GGATCCCGAGGTGAGATTAGTAG 425

RESULT 6  
 LOCUS BF496895 497 bp mRNA EST 19-APR-2001  
 DEFINITION AT10901.5prime AT Drosophila melanogaster adult testes pOTB7  
 Drosophila melanogaster cDNA clone AT10901 5 similar to CG7235:  
 Fban0007235 'heat shock protein' located on: 2L 25F4-25F4;  
 04/08/2001, mRNA sequence.  
 ACCESSION BF496895  
 VERSION BF496895.2 GI:13694366  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 497)  
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nuncio, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Calniker, S., and Rubin, G.M.

TITLE BDGP/HMR AT Drosophila EST Project  
 JOURNAL Unpublished (2000)  
 COMMENT On Dec 6, 2000 this sequence version replaced gi:11580196.  
 Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
 hit genomic AEO03610; arm:2L [5357975,5614606]  
 estimated-cyto:25E2-25F4; 04/08/2001  
 Plate: AT.109 row: A column: 1  
 High quality sequence stop: 455.  
 Location/Qualifiers  
 1..497

FEATURES  
 source  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="AT10901"  
 /sex="male"  
 /dev\_stage="0-3 day old Ore-R males"  
 /lab\_host="Plates AT.10-AT.120: DHS-alpha. Plates AT.121-AT.319: DHS-alpha Tona"  
 /note="Organ: ADULT testes; Vector: pOTB7; Site\_1: EcoRI; Site\_2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."  
 BASE COUNT 125 a 123 c 159 g 90 t

Query Match 70.4%; Score 17.6; DB 11; Length 497;  
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 gatcccccgatgagctaaagctagc 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 452 GATCCCGCGAGGAGTATGCTAGC 475

RESULT 7  
 LOCUS BI038578/c 510 bp mRNA EST 14-JUN-2001  
 DEFINITION MR4-NT0143-220101-004-h09 NT0143 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BI038578  
 VERSION BI038578.1 GI:14445204  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 510)

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&t2=MR4-NT0143-220101-004-h09&t3=2001-01-22&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 196  
 High quality sequence stop: 396.  
 Location/Qualifiers  
 1..510

FEATURES  
 source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NT0143"  
 /dev\_stage="Adult"  
 /note="Organ: nervous tumor; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 144 a 119 c 110 g 137 t  
 ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 510;  
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 ggatcccccgatgagctaaagctag 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 503 GGATCCCGAGGAGTCAATCTTG 480

RESULT 8  
 LOCUS BF877912 587 bp mRNA EST 17-JAN-2001

```

DEFINITION MR0-ET0109-171100-001-a01 ET0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF877912
VERSION 1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 587)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baid,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR0&t2=MR0-ET0109-
171100-001-a01&t3=2000-11-17&t4=1)
171100-001-a01&t3=2000-11-17&t4=1)
High quality sequence start: 19
High quality sequence stop: 521.
FEATURES
source
1..587
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0109"
/dev_stage="Adult"
/notes="Orgn: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196.716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 149 a 131 c 164 g 143 t
ORIGIN
1 ggatcccccgcgtagctaaagctag 24
|||||||
Db 15 GGATCCCCCGATGAGTAATCTTG 38

Query Match 70.4%; Score 17.6; DB 11; Length 587;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggatcccccgcgtagctaaagctag 24
|||||||
Db 15 GGATCCCCCGATGAGTAATCTTG 38

RESULT 9
LOCUS BE639934
DEFINITION 946044B07.y1 946 - tassal primordium prepared by Schmidt lab Zea
mays cDNA, mRNA sequence.
ACCESSION BE639934
VERSION BE639934
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

```

```

REFERENCE 1 (bases 1 to 605)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946044 row: B column: 07.
FEATURES
source
1..605
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassal primordium prepared by Schmidt
lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XLODR"
/notes="Orgn: tassels; Vector: HybriZAP; Site_1: EcoRI;
Site_2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybriZAP. Sample insert size range was 350 bp
to 3 Kb with a 1 Kb average."
BASE COUNT 138 a 160 c 186 g 120 t
ORIGIN
1 ggatcccccgcgtagctaaagctag 25
|||||||
Db 533 GAATGCCCGTGTAGCTAGGCAAGC 557

Query Match 70.4%; Score 17.6; DB 10; Length 605;
Best Local Similarity 80.0%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggatcccccgcgtagctaaagctag 25
|||||||
Db 533 GAATGCCCGTGTAGCTAGGCAAGC 557

RESULT 10
LOCUS A0834638
DEFINITION HS_5442_AL_H01_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1018 Col=1 Row=O, DNA sequence.
ACCESSION A0834638
VERSION A0834638
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 611)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

```

or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
<http://www.htsc.washington.edu>  
 Plate: 1018 row: 0 column: 1  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 611.

#### FEATURES

source

1. .611  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="plate=1018 Col=1 Row=0"  
 /clone\_lib="RPC1-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACe3.6 vector at EcoRI sites"

BASE COUNT 195 a 135 c 105 g 175 t 1 others  
 ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 611;  
 Best Local Similarity 83.3%; Pred. No. 5.5e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggatcccccagtgagctaagctag 24  
 Db 233 GGATCCACCTCTGAGCCAGCTAG 256

#### RESULT 11

CNS038XY/c

LOCUS

DEFINITION

15-MAY-2000

GSS

006608 of library G from Tetraodon nigroviridis, genomic survey

sequence.

AL233215

AL233215.1 GI:7892350

GSS: genome survey sequence.

KEYWORDS

Tetraodon nigroviridis.

ORGANISM

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.

1 (bases 1 to 924)

Roest-Crollius,H., Jallou,O., Dasilva,C., Fizes,C., Fisher,C.,

Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and

Weissenbach,J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Unpublished

2 (bases 1 to 924)

Roest-Crollius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Unpublished

3 (bases 1 to 924)

Genoscope.

Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis

genome. For more information, please take a look at

<http://www.genoscope.cns.fr/Tetraodon>.

Location/Qualifiers

1. .924

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="006G08"

FEATURES

source

#### BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

20; Conservative

0; Mismatches

4; Indels

0; Gaps

0;

QY 1 ggatcccccagtgagctaagctag 24

Db 785 GGCTCCCGCATGAGCAGCTGG 762

RESULT 12

BE120999/c

LOCUS

DEFINITION

UI-R-CA0-baf-h-12-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone

BE120999

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 295)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

normalized hypothalamus library cDNA Library Preparation: M.B.

Soares lab clone distribution: clones will be available through

Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

Location/Qualifiers

1. .295

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-CA0-baf-h-12-0-UI"

/clone\_lib="UI-R-CA0"

/lab\_host="DHI0B (Life Technologies)"

/note="Vector: pWT3D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CA0

library is a subtracted library derived from the following

tissues: thalamus, cerebellum, hypothalamus, medulla, pons

, midbrain, cerebral cortex, corpus striatum, testis, and

hippocampus. For a detailed description of the library

from which this clone was derived, please visit our web

site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been

previously described in (Bonaldo, Lennon and Soares,

Genome Research 6:791-806, 1996)

TAG\_LIB=UI-R-CA0

TAG\_TISSUE=hypothalamus

TAG\_SEQ=CATGC"

#### FEATURES

source

```

BASE COUNT      75 a      67 c      72 g      81 t
ORIGIN

Query Match      68.8%; Score 17.2; DB 10; Length 295;
Best Local Similarity 86.4%; Pred. No. 7.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaagct 22
|||||
Db 237 GGATCCCGACGACGCGAAGCT 216

RESULT 13
BB499943
LOCUS      BB499943      326 bp      mRNA      EST      26-JUL-2000
DEFINITION musculus cDNA clone D630023P14 3', mRNA sequence.
ACCESSION BB499943.1 GI:9476600
VERSION BB499943.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya
T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL:http://genome-gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsumai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source Location/Qualifiers
1. .326
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630023P14"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"

/lab_host="DHI0B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCAACTCGAGTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTTAAATTAATCCCTCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      95 a      68 c      60 g      103 t
ORIGIN

Query Match      68.8%; Score 17.2; DB 10; Length 326;
Best Local Similarity 86.4%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 tcccccgatgagctaagctagc 25
|||||
Db 107 TCCCCGATGAGCGAGGCTGGC 128

RESULT 14
BF410724/c
LOCUS      BF410724      336 bp      mRNA      EST      28-NOV-2000
DEFINITION UI-R-CA0-bmc-b-04-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
UI-R-CA0-bmc-b-04-0-UI 3', mRNA sequence.
ACCESSION BF410724
VERSION BF410724.1 GI:11398699
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
TITLE 1 (bases 1 to 336)
Normalizaton and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dt track served to identify it as a clone from the
normalized thalamus library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers
1. .336
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA0-bmc-b-04-0-UI"
/clone_lib="UI-R-CA0"
/lab_host="DHI0B (Life Technologies)"

```

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAO library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_LIB=UI-R-CAO  
TAG\_TISSUE=thalamus  
TAG\_SEQ=GAATCG"

BASE COUNT 82 a 73 c 85 g 96 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 11; Length 336;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggatcccgatgagctaagct 22  
||||||| |||||  
Db 238 GGATCCCGACGACGAAGCT 217

#### RESULT 15

AQ867402 368 bp DNA GSS 03-NOV-1999  
nbeb0031F05f CUGI Rice BAC Library (ECORI) Oryza sativa genomic  
clone nbeb0031F05f, DNA sequence.

ACCESSION AQ867402  
VERSION AQ867402.1 GI:6217859  
KEYWORDS GSS.

SOURCE Oryza sativa.  
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 368)

Wing,R.A. and Dean,R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)

CONTACT: Wing RA

Clemson University Genomics Institute  
Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: [rwing@clemson.edu](mailto:rwing@clemson.edu)

Seq primer: TAATACGACTCACTATAGGG

Class: BAC ends

High quality sequence start: 23

High quality sequence stop: 328.

Location/Qualifiers

1..368

/Organism="Oryza sativa"

/Strain="Japonica"

/cultivar="Nipponbare"

/db\_xref="taxon:4530"

/clone\_lib="nbeb0031F05f"

/clone\_lib="CUGI Rice BAC Library (ECORI)"

/tissue\_type="Leaf"

/lab\_host="E. coli DH10B"

/note="Vector: pBACindigo; Site\_1: EcoRI; Site\_2: EcoRI;  
Rice is the most important food crop in the world. Half of  
the world population, especially those inhabiting highly  
populated areas of the humid tropics and subtropics, rely  
on rice as their primary source of carbohydrate.  
Monocotyledonous rice is a diploid plant (2n=24) with a  
haploid genome equivalent of 431 Mbp (Arumuganathan and  
Earle, 1991). The relatively small genome of rice, three  
times larger than that of Arabidopsis, makes it suitable

for genomic studies. In order to facilitate positional  
cloning, physical mapping and genome sequencing of rice,  
we have constructed a BAC library from Oryza sativa,  
Nipponbare variety using EcoRI as the cloning enzyme. The  
library contains 55,296 clones with an average insert size  
of 121 kb providing approximately 15 haploid genome  
equivalents. The deep coverage allows the isolation a  
particular sequence with a probability of 99.9 %. Three  
high density filters, each containing 18,432 clones  
(doubly spotted), represent the whole library for colony  
screening and can be requested from the Clemson University  
BAC/EST Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu))."

BASE COUNT 101 a 85 c 78 g 104 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 13; Length 368;  
Best Local Similarity 86.4%; Pred. No. 7.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atcccccgatgagctaagctag 24  
||| ||| ||||| |||||  
Db 34 ATCACCACATGAGCTCAGCTAG 55

Search completed: February 15, 2002, 18:03:51

Job time: 20940 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:47:05 ; Search time 2553.1 Seconds  
(without alignments)  
142.156 Million cell updates/sec

Title: US-09-698-903B-6

Perfect score: 22  
Sequence: 1 tcatctacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Match	Length	DB	ID	Description
1	22	100.0	22	6	AX127753	AX127753 Sequence
2	22	100.0	22	6	AX172462	AX172462 Sequence
3	22	100.0	249	12	ARGMTUB	X05579 Soybean bet
4	22	100.0	831	1	ATTRN7	X00091 Agrobacteri
5	22	100.0	878	1	ATTN7	X00431 Agrobacteri
6	22	100.0	1037	6	AX10942	AX10942 Nucleotide
7	22	100.0	1085	6	AX10939	AX10939 Nucleotide
8	22	100.0	1160	6	AX10943	AX10943 Nucleotide
9	22	100.0	1166	6	AX10941	AX10941 Nucleotide
10	22	100.0	1186	6	AX18051	AX18051 DNA used as
11	22	100.0	1186	6	AR095107	AR095107 Sequence
12	22	100.0	1186	6	AR098313	AR098313 Sequence
13	22	100.0	1186	6	AX012338	AX012338 Sequence
14	22	100.0	1186	6	I49886	I49886 Sequence 2
15	22	100.0	1186	6	I82374	I82374 Sequence 2
16	22	100.0	2476	12	TBI251013	AJ251013 Transform
17	22	100.0	3201	6	I44104	I44104 Sequence 23
18	22	100.0	3201	6	I44103	I44103 Sequence 22
19	22	100.0	3236	12	TBI251014	AJ251014 Transform
20	22	100.0	4832	6	AX172441	AX172441 Sequence
21	22	100.0	4946	6	A60108	A60108 Sequence 1
22	22	100.0	4946	6	A76915	A76915 Sequence 1
23	22	100.0	4946	6	AR098307	AR098307 Sequence
24	22	100.0	4946	6	AX172440	AX172440 Sequence
25	22	100.0	5349	6	A71437	A71437 Sequence 7
26	22	100.0	5560	6	A60112	A60112 Sequence 5
27	22	100.0	5560	6	AR098311	AR098311 Sequence
28	22	100.0	5865	6	AX127748	AX127748 Sequence
29	22	100.0	5865	6	AX127748	AX127748 Sequence
30	22	100.0	6539	6	E31991	E31991 Mutated bar
31	22	100.0	6548	6	A60109	A60109 Sequence 2
32	22	100.0	6548	6	A76916	A76916 Sequence 2
33	22	100.0	6548	6	AR098308	AR098308 Sequence
34	22	100.0	6548	6	E31990	E31990 Mutated bar
35	22	100.0	7566	6	A24783	A24783 plasmid pPS
36	22	100.0	7566	6	AR074388	AR074388 Sequence
37	22	100.0	7599	6	AX063413	AX063413 Sequence
38	22	100.0	7639	6	A24782	A24782 plasmid pJD
39	22	100.0	7639	6	AR074387	AR074387 Sequence
40	22	100.0	7811	6	AR078675	AR078675 Sequence
41	22	100.0	12095	12	BINHYGDNA	Z37515 Binary vect
42	22	100.0	24595	1	ATACH5	X00493 Agrobacteri
43	22	100.0	24595	6	E00404	E00404 Ti plasmid
44	22	100.0	24595	6	E00546	E00546 DNA fragmen
45	22	100.0	194140	1	AF242881	AF242881 Agrobacte

ALIGNMENTS

RESULT 1

AX127753	AX127753	Sequence 6 from Patent WO0131042.	22 bp	DNA	PAT	15-MAY-2001
LOCUS	AX127753	Sequence 6 from Patent WO0131042.	22 bp	DNA		
DEFINITION	AX127753	Sequence 6 from Patent WO0131042.	22 bp	DNA		
ACCESSION	AX127753	Sequence 6 from Patent WO0131042.	22 bp	DNA		
VERSION	AX127753.1	GI:14134400				
KEYWORDS		synthetic construct.				
SOURCE		synthetic construct				
ORGANISM		artificial sequence.				

REFERENCE 1 (bases 1 to 22)  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 6 03-MAY-2001;  
Aventis CropScience N.V. (BE)

FEATURES  
source  
1..22  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer MDB193"

```

BASE COUNT      6 a      7 c      4 g      5 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 22; DB 6; Length 22;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22
|||||
Db 1 TCATCTACGGCAATGTACCAGC 22

RESULT 2
AXI72462
LOCUS AXI72462 22 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 23 from Patent WO0141558.
ACCESSION AXI72462
VERSION AXI72462.1 GI:14597574
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
          artificial sequence.
REFERENCE 1 (bases 1 to 22)
AUTHORS de Both,G. and de Beuckeleer,M.
TITLE Hybrid winter oilseed rape and methods for producing same
JOURNAL Patent: WO 0141558-A 23 14-JUN-2001;
FEATURES
    source
        Location/Qualifiers
            1..22
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                /note="primer 193"
BASE COUNT      6 a      7 c      4 g      5 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 22; DB 6; Length 22;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22
|||||
Db 1 TCATCTACGGCAATGTACCAGC 22

RESULT 3
ARGMTUB
LOCUS ARGMTUB 249 bp DNA SYN 02-APR-1988
DEFINITION Soybean beta-1-tubulin gene fused to T1 plasmid unit 7 3UTR.
ACCESSION X05579
VERSION X05579.1 GI:59087
KEYWORDS beta-tubulin; fusion gene; plasmid.
SOURCE synthetic construct.
ORGANISM synthetic construct
          artificial sequence.
REFERENCE 1 (bases 1 to 249)
AUTHORS Guiltinan,M.J., Velten,J., Bustos,M.M., Cyr,R.J., Schell,J. and
          Fosket,D.E.
TITLE The expression of a chimeric soybean beta-tubulin gene in tobacco
JOURNAL Mol. Gen. Genet. 207, 328-334 (1987)
FEATURES
    source
        Location/Qualifiers
            1..249
                /organism="synthetic construct"
                /db_xref="taxon:32630"
            1..51
                /note="fusion product (17AA); Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA29084.1"
                /db_xref="GI:4376141"
                /translation="AMAAWASSNWSSTDPMS"
CDS

```

---

```

misc_feature 1..13
              /note="beta-1-tubulin sequence"
misc_feature 14..28
              /note="pUC 13 polylinker"
misc_feature 29..35
              /note="Sal I linker"
misc_feature 37
              /note="theroretical fusion junction (24) with gene 7 of T1
              plasmid"
misc_feature 173..178
              /note="put.polyA signal"
polyA_site 198
            /note="polyA site"
misc_feature 220..225
            /note="put.polyA signal"
BASE COUNT    76 a    47 c    38 g    88 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 22; DB 12; Length 249;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22
|||||
Db 106 TCATCTACGGCAATGTACCAGC 127

RESULT 4
ATTRN7
LOCUS ATTRN7 831 bp DNA BCT 02-SEP-1999
DEFINITION Agrobacterium tumefaciens gene encoding 'transcript 7' mRNA for a
          protein with unknown function.
ACCESSION V00090
VERSION V00090.1 GI:39180
KEYWORDS unidentified reading frame.
SOURCE Agrobacterium tumefaciens.
ORGANISM Agrobacterium tumefaciens
          Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
          Rhizobiaceae; Rhizobium.
REFERENCE 1 (bases 1 to 831)
AUTHORS Dhaese,P., De Greve,H., Gielen,J., Seurlinck,J., Van Montagu,M.M.
          and Schell,J.
TITLE Identification of sequences involved in the polyadenylation of
          higher plant nuclear transcripts using Agrobacterium T-DNA genes as
          models
JOURNAL EMBO J. 2, 419-426 (1983)
REFERENCE 2 (bases 76 to 100)
AUTHORS Dhaese,P.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-1983) to the EMBL/GenBank/DBJ databases
COMMENT Data kindly reviewed (27-MAY-1983) by Dhaese P.
FEATURES
    Location/Qualifiers
        1..831
            /organism="Agrobacterium tumefaciens"
            /strain="(octopine TL-DNA)"
            /db_xref="taxon:358"
            75..81
                /note="CG [1] revised CCAGAGG [2]"
                /citation=1
            /citation=2
            99..101
                /note="CCA [1] revised CTA [2]"
                /citation=1
            /citation=2
            132..673
                /note="transcript 7 (alternate)"
            132..672
                /note="transcript 7"
            148..528
                /note="unknown gene (148 is 1st base in codon) (525 is 3rd
                base in codon)"
            /codon_start=1

```

/transl\_table=11  
/protein\_id="CAA23429.1"  
/db\_xref="GI:39181"  
/db\_xref="SWISS-PROT:P03867"  
/translation="MNFADTPLASLDLDWACEEFIKTYGASPOLGTGEVIQTNNGLLY  
LYGKSLSORLHDTHLKFKEKEELSFTTIKPAEMKAQOSDLTYVVAIQSNYFLCVSN  
PEKGLFCHNRPFLYPIVAGHSGMS"  
BASE COUNT 262 a 176 c 138 g 255 t  
ORIGIN

Query Match 100.0%; Score 22; DB 1; Length 831;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||  
Db 580 TCATCTACGGCAATGTACCAGC 601

RESULT 5  
ATTDNA  
LOCUS Agrobacterium tumefaciens crown gall tumor T-DNA from Ti (tumor  
DEFINITION inducing) plasmid pTiA6. 25-MAR-1996  
ACCESSION X00431  
VERSION X00431.1 GI:39150  
KEYWORDS plasmid.  
SOURCE Agrobacterium tumefaciens.  
ORGANISM Agrobacterium tumefaciens.  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Rhizobiaceae; Rhizobium.  
REFERENCE 1 (bases 1 to 878)  
AUTHORS McPherson,J.C.  
TITLE DNA sequence analysis of crown gall tumor T-DNA encoding the 0.7 kb  
transcript  
JOURNAL Nucleic Acids Res. 12 (5), 2317-2325 (1984)  
MEDLINE 84169535  
FEATURES  
source Location/Qualifiers  
1..878  
/organism="Agrobacterium tumefaciens"  
/strain="plasmid pTiA6"  
/db\_xref="taxon:358"  
60..66  
/note="TATA-box"  
68..613  
/note="polyadenylation signal"  
109..489  
/note="unidentified reading frame"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA25129.1"  
/db\_xref="GI:39151"  
/db\_xref="SWISS-PROT:P03867"  
/translation="MNFADTPLASLDLDWACEEFIKTYGASPOLGTGEVIQTNNGLLY  
LYGKSLSORLHDTHLKFKEKEELSFTTIKPAEMKAQOSDLTYVVAIQSNYFLCVSN  
PEKGLFCHNRPFLYPIVAGHSGMS"  
659..664  
misc\_feature  
/note="polyadenylation signal"  
BASE COUNT 288 a 189 c 139 g 262 t  
ORIGIN

Query Match 100.0%; Score 22; DB 1; Length 878;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||  
Db 541 TCATCTACGGCAATGTACCAGC 562

RESULT 6

Al0942  
LOCUS Al0942 1037 bp DNA PAT 27-SEP-1993  
DEFINITION Nucleotide sequence 4 from patent number DE3920034.  
ACCESSION Al0942  
VERSION Al0942.1 GI:492369  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1037)  
AUTHORS  
JOURNAL Patent: DE 3920034-A 4 31-MAY-1990;  
FEATURES Location/Qualifiers  
1..1037  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 338 a 174 c 166 g 359 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1037;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||  
Db 837 TCATCTACGGCAATGTACCAGC 858

RESULT 7  
Al0939  
LOCUS Al0939 1085 bp DNA PAT 27-SEP-1993  
DEFINITION Nucleotide sequence 1 from patent number DE3920034.  
ACCESSION Al0939  
VERSION Al0939.1 GI:492367  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1085)  
AUTHORS  
JOURNAL Patent: DE 3920034-A 1 31-MAY-1990;  
FEATURES Location/Qualifiers  
1..1085  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 369 a 218 c 155 g 343 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1085;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||  
Db 885 TCATCTACGGCAATGTACCAGC 906

RESULT 8  
Al0943  
LOCUS Al0943 1160 bp DNA PAT 27-SEP-1993  
DEFINITION Nucleotide sequence 5 from patent number DE3920034.  
ACCESSION Al0943  
VERSION Al0943.1 GI:492370  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1160)  
AUTHORS  
JOURNAL Patent: DE 3920034-A 5 31-MAY-1990;  
FEATURES Location/Qualifiers

```
source 1..1160
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 367 a 194 c 188 g 411 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1160;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattctacggcaatgtaccagc 22
|||||
Db 960 TCATCTACGGCAATGTACCAGC 981

RESULT 9
LOCUS A10941 1166 bp DNA 27-SEP-1993
DEFINITION Nucleotide sequence 3 from patent number DE3920034.
ACCESSION A10941
VERSION A10941.1 GI:492368
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1166)
AUTHORS
JOURNAL
FEATURES
source
BASE COUNT 381 a 208 c 196 g 381 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1166;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattctacggcaatgtaccagc 22
|||||
Db 1066 TCATCTACGGCAATGTACCAGC 1087

RESULT 10
LOCUS A18051 1186 bp DNA 26-JUL-1994
DEFINITION DNA used as a probe for neo gene seq ID No:2.
ACCESSION A18051
VERSION A18051.1 GI:583120
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1186)
AUTHORS
JOURNAL
FEATURES
source
promoter
gene
CDS

/organism="synthetic construct"
/db_xref="taxon:32630"
1..8
/note="sequence derived from tapetum specific promoter of
Nicotiana tabacum"
167..790
/gene="neomycine phosphotransferase gene"
167..790
/gene="neomycine phosphotransferase gene"
/note="Protein sequence is in conflict with the conceptual
translation"
/codon_start=1

/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 367 a 194 c 188 g 411 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattctacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 11
LOCUS AR095107 1186 bp DNA 08-SEP-2000
DEFINITION Sequence 2 from patent US 6002070.
ACCESSION AR095107
VERSION AR095107.1 GI:10022665
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1186)
AUTHORS
D'Halluin,K. and Gobel,E.
TITLE
Process for transforming monocotyledonous plants
JOURNAL
Patent: US 6002070-A 2 14-DEC-1999;
FEATURES
Location/Qualifiers
1..1186
source
BASE COUNT 244 a 317 c 325 g 300 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattctacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 12
LOCUS AR098313 1186 bp DNA 14-FEB-2001
DEFINITION Sequence 2 from patent US 6074877.
ACCESSION AR098313
VERSION AR098313.1 GI:12807570
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1186)
AUTHORS
D'Halluin,K. and Gobel,E.
TITLE
Process for transforming monocotyledonous plants
JOURNAL
Patent: US 6074877-A 2 13-JUN-2000;
FEATURES
Location/Qualifiers
1..1186
source
BASE COUNT 244 a 317 c 325 g 300 t
ORIGIN

/note="3' regulatory sequence containing the
polyadenylation site derived from agrobacterium T-DNA gene
7"
```

Query Match 100.0%; Score 22; DB 6; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22  
|||||  
Db 1029 TCATCTACGGCAATGTACCAGC 1050

## RESULT 13

AX012338 AX012338 1186 bp DNA PAT 06-SEP-2000

LOCUS Sequence 2 from Patent EP0955371.

AX012338

VERSION AX012338.1 GI:9998387

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 1186)

AUTHORS D'Halluin,K. and Goebel,E.D.

TITLE Process for transforming monocotyledonous plants

JOURNAL Patent: EP 0955371-A 2 10-NOV-1999;

PLANT GENETIC SYSTEMS NV (BE)

FEATURES Location/Qualifiers

source 1..1186

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="DNA used as probe for neo gene"

misc\_feature 1..8

/note="sequence derived from tapetum specific promoter of

Nicotiana tabacum"

misc\_feature 9..790

/note="coding sequence of neomycin phosphotransferase"

misc\_feature 791..1186

/note="3' regulatory sequence containing the

polyadenylation site derived from Agrobacterium T-DNA gene

7"

BASE COUNT 244 a 317 c 325 g 300 t

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22  
|||||  
Db 1029 TCATCTACGGCAATGTACCAGC 1050

## RESULT 14

I49886 I49886 1186 bp DNA PAT 07-OCT-1997

LOCUS Sequence 2 from patent US 5641664.

AX012338

VERSION I49886.1 GI:2472106

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1186)

AUTHORS D'Halluin,K. and Goebel,E.

TITLE Process for transforming monocotyledonous plants

JOURNAL Patent: US 5641664-A 2 24-JUN-1997;

FEATURES Location/Qualifiers

source 1..1186

/organism="unknown"

BASE COUNT 244 a 317 c 325 g 300 t

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22  
|||||  
Db 1029 TCATCTACGGCAATGTACCAGC 1050

## RESULT 15

I82374

LOCUS I82374 1186 bp DNA PAT 10-JUN-1998

DEFINITION Sequence 2 from patent US 5712135.

AX012338

VERSION I82374.1 GI:3210671

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1186)

AUTHORS D'Halluin,K. and Gobel,E.

TITLE Process for transforming monocotyledonous plants

JOURNAL Patent: US 5712135-A 2 27-JAN-1998;

FEATURES Location/Qualifiers

source 1..1186

/organism="unknown"

BASE COUNT 244 a 317 c 325 g 300 t

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22  
|||||  
Db 1029 TCATCTACGGCAATGTACCAGC 1050

Search completed: February 15, 2002, 18:47:06  
Job time: 20065 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:01:18 ; Search time 868.33 Seconds  
(without alignments)  
21.721 Million cell updates/sec

Title: US-09-698-903b-6

Perfect score: 22

Sequence: 1 tcattacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_ll01:\*

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	AAH25438	PCR primer for the
2	22	100.0	22	AAO06995	PCR primer MDB193
3	22	100.0	1037	AAO04705	USP-Promoter-casse
4	22	100.0	1085	AAO04703	Legumin-signalpept
5	22	100.0	1160	AAO04706	USP-signalpeptide
6	22	100.0	1166	AAO04704	USP-Promoter-casse
7	22	100.0	1186	AAO25707	Chimeric neo gene
8	22	100.0	1303	AAAT39337	Plasmid pT588 (Eco
9	22	100.0	3153	AAZ29122	Plasmid DV131 comp
10	22	100.0	3201	AAQ14529	pPS029 Bt ICP codi
11	22	100.0	3201	AAQ15144	pVE36 Bt ICP codin

Plasmid DV130 comp

Plasmid DV133 used

Plasmid DV132 used

Nucleotide sequenc

T-DNA of plasmid p

Nucleotide sequenc

Plasmid pT5172delt

T-DNA of pTTS24.

Plasmid pTCC0113 T-

Chimeric T-DNA of

Chimeric T-DNA of

Plasmid pT5174 use

Plasmid pT5172. C

E. coli plasmid pT

Plasmid pT5346. U

Plasmid pPS0212 co

Nucleotide sequenc

Plasmid pJD884 con

Sequence of opine

Complete nucleotid

T-DNA right border

Transformed Arabid

PCR primer for the

PCR primer MDB258

PCR primer T4 used

Right (5') border

Left (3') border f

Right flanking reg

Fusarium venenatum

Human liver GPI-PL

Glycosyl-phosphati

Polynucleotide seq

ALIGNMENTS

RESULT 1

AAH25438 1

ID AAH25438 standard; DNA; 22 BP.

XX AC AAH25438;

XX DT 22-AUG-2001 (first entry)

XX DE PCR primer for the right flanking region in transgenic plant BN-RFL.

XX KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;

XX KW fertility restorer gene; barnase gene; barstar gene; ss.

XX OS Synthetic.

XX PN WO200141558-A1.

XX PD 14-JUN-2001.

XX PF 06-DEC-2000; 2000WO-EP12872.

XX PR 08-DEC-1999; 99US-0457037.

XX PA (AVET ) AVENTIS CROPS SCIENCE NV.

XX PI De Both G, De Beuckeleer M;

XX DR WPI; 2001-381419/40.

XX PT Transgenic winter oilseed rape plants suited for producing hybrid seed

XX PT with improved qualities, comprises a male-sterility gene and fertility

XX PS restorer gene, integrated into the genome -

PS Claim 60; Page 46; 98pp; English.

The specification describes a pair of transgenic winter oilseed rape plants suited for producing hybrid seed. One of the plants has an expression cassette comprising a male-sterility gene (e.g. barnase gene), and the other plant has an expression cassette comprising a fertility restorer gene (e.g. barstar gene), integrated into the genome. The fertility restorer gene is capable of preventing the activity of the male-sterility gene. The plant pair is useful for producing hybrid seed. Plants developed from the hybrid seed have agronomic performance, genetic stability and adaptability to different genetic backgrounds. The present PCR primer was used to amplify the right flanking region of a vector in a transgenic plant which carries the TA29-barstar transgene.

Sequence 22 BP; 6 A; 7 C; 4 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.034;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 ||||||||||||||||||  
 Db 1 tcattctacggcaatgtaccagc 22

RESULT 2  
 AAD06995  
 ID AAD06995 standard; DNA; 22 BP.  
 AC AAD06995;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE PCR primer MBI193 to generate the flanking region of elite event MS-B2.  
 XX  
 KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; thermal asymmetric interlaced;  
 KW TAIL; ss.  
 XX  
 OS Agrobacterium sp.  
 XX  
 PN WO200131042-A2.  
 XX  
 PD 03-MAY-2001.  
 XX  
 PF 26-OCT-2000; 2000WO-EP10680.  
 XX  
 PR 29-OCT-1999; 99US-0430497.  
 XX  
 PA (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 PI Weston B, De Beuckeleer M;  
 XX  
 DR WPI; 2001-300517/31.  
 XX  
 PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX  
 PS Example 3; Page 28; 53pp; English.  
 XX  
 CC The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is secondary thermal interlaced (TAIL)-PCR primer  
 CC MBI193 used to right (5') border flanking region of elite event MS-B2.  
 CC This primer corresponds to position 226-247 of plasmid pRCO113.

SQ Sequence 22 BP; 6 A; 7 C; 4 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.034;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 ||||||||||||||||||  
 Db 1 tcattctacggcaatgtaccagc 22

RESULT 3  
 AAQ04705  
 ID AAQ04705 standard; DNA; 1037 BP.  
 XX  
 AC AAQ04705;  
 XX  
 DT 12-OCT-1990 (first entry)  
 XX  
 DE USP-Promoter-cassette USP-Pr.T7.1.  
 XX  
 KW Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.  
 XX  
 PN DE3920034-A.  
 XX  
 PD 31-MAY-1990.  
 XX  
 PF 20-JUN-1989; 89DE-3920034.  
 XX  
 PR 19-SEP-1988; 88DD-0319887.  
 XX  
 PA (PFLA-) VE KOMB PFLANZENZUC.  
 XX  
 PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;  
 XX  
 DR WPI; 1990-172459/23.  
 XX  
 PT Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.  
 XX  
 PS Disclosure; ; pp; German.  
 XX  
 CC The unique BglII-Ort (720-725) site is for ligating foreign DNA and the  
 CC HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 CC cassette in the Ti-vector pGA471. Agrobacterium tumefaciens is  
 CC transfected.  
 CC See also AAQ04703-Q04706.  
 XX  
 SQ Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 100.0%; Score 22; DB 11; Length 1037;  
 Best Local Similarity 100.0%; Pred. No. 0.059;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 ||||||||||||||||||  
 Db 837 tcattctacggcaatgtaccagc 858

RESULT 4  
 AAQ04703  
 ID AAQ04703 standard; DNA; 1085 BP.  
 XX  
 AC AAQ04703;  
 XX  
 DT 12-OCT-1990 (first entry)  
 XX  
 DE Legumin-signalpeptide cassette Le-Sig.T7.  
 XX  
 KW Foreign DNA incorporation; recombinant DNA techniques;

higher plant genome; legumin; B4 gene; signalpeptide; Le-Sig.T7.; ss.

Key Location/Qualifiers  
CDS 747..814  
/\*tag= a  
/product=Legumin-signalpeptide

DE3920034-A.

31-MAY-1990.

20-JUN-1989; 89DE-3920034.

19-SEP-1988; 88DD-0319887.

(PFLA-) VE KOMB PFLANZENZUC.

Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

WPI; 1990-172459/23.

P-PSDB; AAR05198.

Incorporation of DNA into higher plant genome - by specified  
recombinant DNA techniques.

Disclosure; ; pp; German.

The Legumin gene B4 is used. The unique BglII-Ort (815-820) site is for  
ligating foreign DNA and the HindIII-Ort in the 3' polylinker  
(1080-1085) for cloning the cassette in the Ti-vector pGA471.  
The cassette is cloned into the binary Ti-vectors pGA471 and  
Agrobacterium tumefaciens is transfected.  
See also AAQ04703-Q04706.

Sequence 1085 BP; 368 A; 219 C; 155 G; 343 T; 0 other;

Query Match 100.0%; Score 22; DB 11; Length 1085;  
Best Local Similarity 100.0%; Pred. No. 0.059;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
|||||  
Db 885 tcattctacggcaatgtaccagc 906

RESULT 5

AAQ04706  
ID AAQ04706 standard; DNA; 1160 BP.

AC AAQ04706;

12-OCT-1990 (first entry)

USP-signalpeptide cassette USP-Sig.T7.

Foreign DNA incorporation; recombinant DNA techniques;  
higher plant genome; signalpeptide; USP-Sig.T7.; ss.

Key Location/Qualifiers  
CDS 708..877  
/\*tag= a  
/product=signalpeptide  
747..817

Intron

DE3920034-A.

31-MAY-1990.

20-JUN-1989; 89DE-3920034.

19-SEP-1988; 88DD-0319887.

(PFLA-) VE KOMB PFLANZENZUC.

Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

WPI; 1990-172459/23.

P-PSDB; AAR05199.

Incorporation of DNA into higher plant genome - by specified  
recombinant DNA techniques.

Disclosure; ; pp; German.

The unique BglII-Ort (890-895) site is for  
ligating foreign DNA and the HindIII-Ort in the 3' polylinker  
(1155-1160) for cloning the cassette in the Ti-vector pGA471.  
The cassette is cloned into the binary Ti-vectors pGA471 and  
Agrobacterium tumefaciens is transfected.  
See also AAQ04703-Q04706.

Sequence 1160 BP; 369 A; 192 C; 188 G; 411 T; 0 other;

Query Match 100.0%; Score 22; DB 11; Length 1160;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
|||||  
Db 960 tcattctacggcaatgtaccagc 981

RESULT 6

AAQ04704  
ID AAQ04704 standard; DNA; 1166 BP.

AC AAQ04704;

12-OCT-1990 (first entry)

USP-Promoter-cassette USP-Pr.T7.2.

Foreign DNA incorporation; recombinant DNA techniques;  
higher plant genome; legumin; USP-Pr.T7-2; ss.

DE3920034-A.

31-MAY-1990.

20-JUN-1989; 89DE-3920034.

19-SEP-1988; 88DD-0319887.

(PFLA-) VE KOMB PFLANZENZUC.

Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

WPI; 1990-172459/23.

Incorporation of DNA into higher plant genome - by specified  
recombinant DNA techniques.

Disclosure; ; pp; German.

The unique BglII-Ort (720-725) site is for ligating foreign DNA and the  
HindIII-Ort in the 3' polylinker (1261-1266) for cloning the  
cassette in the Ti-vector pGA471. Agrobacterium tumefaciens is  
transfected.  
See also AAQ04703-Q04706.

Sequence 1166 BP; 381 A; 208 C; 196 G; 381 T; 0 other;

Query Match 100.0%; Score 22; DB 11; Length 1166;

Best Local Similarity 100.0%; Pred. No. 0.06; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcactcagcgcaatgtaccagc 22  
 Db 1066 tcactcagcgcaatgtaccagc 1087

```

RESULT 7
AAQ25707
ID AAQ25707 standard; DNA; 1186 BP.
XX
AC AAQ25707;
XX
DT 07-DEC-1992 (first entry)
DE Chimeric neo gene probe.
XX
KW Nicotiana tabacum; neomycine phosphotransferase; Agrobacterium; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT promoter 1..8
FT /tag= a
FT /note= "sequence derived from tapetum specific
FT promoter of Nicotiana tabacum"
FT
FT CDS 9..790
FT /tag= b
FT /product= neomycine_phosphotransferase
FT /tag= c
FT /note= "3' regulatory sequence contg. the
FT polyadenylation site derived from
FT Agrobacterium T-DNA gene 7"
XX
PN WO9209696-A.
XX
PD 11-JUN-1992.
XX
PF 21-NOV-1991; 91WO-EP02198.
XX
PR 23-NOV-1990; 90EP-0403332.
PR 08-JUL-1991; 91EP-0401888.
XX
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
PI Dhalluin K, Goebel E;
XX
DR WPI; 1992-217075/26.
XX
PT Transforming monocotyledonous plants e.g. cereals - comprises
PT wounding and/or degrading cells of intact plant tissue or
PT embryogenic callus
XX
PS Disclosure; Page 60; 76pp; English.
XX
CC Two transformed corn plants were analysed by means of Southern
CC hybridisation. As a probe, a 1184 bp EcoRI-HindIII fragment derived
CC from another plasmid was used. The sequence of that plasmid is
CC given below. Results showed that at least a chimeric neo gene was
CC integrated into the plant genomic DNA.
XX
SQ Sequence 1186 BP; 244 A; 317 C; 325 G; 300 T; 0 other;

```

Query Match 100.0%; Score 22; DB 13; Length 1186;  
 Best Local Similarity 100.0%; Pred. No. 0.06;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcactcagcgcaatgtaccagc 22  
 Db 1029 tcactcagcgcaatgtaccagc 1050

```

RESULT 8
AAT39337
ID AAT39337 standard; DNA; 1303 BP.
XX
AC AAT39337;
XX
DT 22-JAN-1997 (first entry)
DE Plasmid pTS88 (EcoRI-HindIII fragment).
XX
KW Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1..35
FT /tag= a
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
FT promoter 36..694
FT /tag= b
FT /label= p35S
FT /function= 35S promoter of cauliflower mosaic virus
FT strain CM1841
FT CDS 695..967
FT /tag= c
FT /label= barstar
FT /product= Bacillus amyloliquefaciens barstar
FT polyA_signal 968..1287
FT /tag= d
FT /label= 3'g7
FT /function= region containing polyadenylation signal
FT of gene 7 og Agrobacterium T-DNA
FT misc_feature 1288..1303
FT /tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX
WO9626283-Al.
XX
PD 29-AUG-1996.
XX
PF 21-FEB-1996; 96WO-EP00722.
XX
PR 21-FEB-1995; 95EP-0400364.
XX
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
PI Botterman J, Cornelissen M, Michiels F;
XX
DR WPI; 1996-402373/40.
XX
PT Prodn. of male sterile plants by transforming with a chimaeric
PT construct - comprising a male sterility DNA e.g. barnase and a
PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
PT for generating hybrid cultivars
XX
PS Example 1; Page 38; 56pp; English.
XX
CC The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains
CC barstar DNA under control of a 35S promoter. The plasmid was
CC used with pTS174 (see also AAT39336) contg. barnase DNA under
CC control of the stamen-specific promoter EI to produce male sterile
CC rice cv. Kochihibiki transgenic plants, and with plasmid pV5136
CC (see also AAT39338) contg. barnase DNA under control of the stamen-
CC specific PCA55 promoter to produce male sterile maize plants.
CC Expression of barnase (a ribonuclease) in the stamen leads to male
CC sterility. Constitutive expression of barstar counteracts possible
CC low level expression of barnase DNA in non-stamen tissue.
XX

```

SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 100.0%; Score 22; DB 17; Length 1303;  
 Best Local Similarity 100.0%; Pred. No. 0.061;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 |||||  
 Db 1094 tcattctacggcaatgtaccagc 1115

## RESULT 9

AAZ29122  
 ID AAZ29122 standard; DNA; 3153 BP.

XX  
 AC AAZ29122;

XX  
 DT 21-FEB-2000 (first entry)

XX Plasmid DV131 comprising L3/EGFP:NPTII/Tr7 expression cassette.

XX Transgenic seed; marker: aleurone-specific promoter; Plasmid DV131;  
 KW GFP:NPTII fusion protein construct; L3/EGFP:NPTII/Tr7; maize R gene;  
 KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
 KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene; assay;  
 KW Green fluorescent protein; GFP; gene fusion; selection; screening;  
 KW expression; automated seed screening technique; screenable marker;  
 KW transformant; embryogenic tissue; implementation; ds.

XX  
 OS Synthetic.

XX  
 PN WO9960129-A1.

XX  
 PD 25-NOV-1999.

XX  
 PF 18-MAY-1999; 99WO-US11023.

XX  
 PR 18-MAY-1998; 98US-0080625.

XX  
 PA (DEKA-) DEKALB GENETICS CORP.

XX  
 PI Kriz AL, Spencer TM;

XX  
 DR WPI; 2000-072441/06.

XX Screenable marker genes useful for identification of transgenic seeds  
 for plant breeding -

XX  
 PS Example 1; Page 164-166; 182pp; English.

XX The present DNA sequence is the plasmid DV131, that is used in the  
 CC generation of GFP:NPTII fusion protein constructs. It contains an  
 CC expression cassette comprising, a promoter from the maize L3 oleosin  
 CC gene, the coding sequence of EGFP:NPTII translational fusion, excised  
 CC from DV126 and the Tr7 terminator. This plasmid is used to carry a  
 CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to the gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.

XX Sequence 3153 BP; 801 A; 857 C; 756 G; 739 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3153;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 |||||  
 Db 2675 tcattctacggcaatgtaccagc 2696

## RESULT 10

AAQ14529  
 ID AAQ14529 standard; DNA; 3201 BP.

XX  
 AC AAQ14529;

XX  
 DT 27-JAN-1992 (first entry)

XX pPS029 Bt ICP coding sequence.

XX  
 KW Bacillus thuringiensis; insecticidal crystal protein; ICP;  
 KW deletion; ss.

XX  
 OS Synthetic.

XX  
 PN WO9116432-A.

XX  
 PD 31-OCT-1991.

XX  
 PF 17-APR-1991; 91WO-EP00733.

XX  
 PR 18-APR-1990; 90EP-0401055.

XX  
 PA (PLAN-) PLANT GENETIC SYST.

XX  
 PI Cornelissen M, Soetaert P, Stam M, Dockx J;

XX  
 DR WPI; 1991-339820/46.

XX Modified Bacillus thuringiensis insecticidal crystal protein  
 PT genes - having A and T sequences changed to G and C sequences  
 PT encoding same amino acids, for increased expression levels

XX  
 PS Disclosure; Fig 6(c); 78pp; English.

XX  
 CC "n" in the sequence refers to not known nucleotides.

XX pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-  
 CC terminal modification and the internal modification of the Bt ICP  
 CC coding sequence.

XX  
 CC See also AAQ14529, AAQ15142-44.

XX  
 SQ Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;

## Query Match

Best Local Similarity 100.0%; Score 22; DB 12; Length 3201;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattctacggcaatgtaccagc 22  
 |||||  
 Db 2948 tcattctacggcaatgtaccagc 2969

## RESULT 11

AAQ15144  
 ID AAQ15144 standard; DNA; 3201 BP.

XX  
 AC AAQ15144;

XX  
 DT 27-JAN-1992 (first entry)

XX pVE36 Bt ICP coding sequence.

XX  
 KW Bacillus thuringiensis; insecticidal crystal protein; ICP;  
 KW deletion; ss.

XX  
 OS Synthetic.

XX PN W09116432-A.  
 XX PD 31-OCT-1991.  
 XX PF 17-APR-1991; 91WO-EP00733.  
 XX PR 18-APR-1990; 90EP-0401055.  
 XX PA (PLAN-) PLANT GENETIC SYST.  
 XX PI Cornelissen M, Sostaert P, Stam M, Dockx J;  
 XX DR WPI; 1991-339820/46.  
 XX PT Modified *Bacillus thuringiensis* insecticidal crystal protein  
 PT genes - having A and T sequences changed to G and C sequences  
 PT encoding same amino acids, for increased expression levels  
 XX PS Disclosure; Fig 6(c); 78pp; English.  
 XX CC "n" in the sequence refers to not known nucleotides.  
 CC pPS029 (AAQ14529) is identical to pVE36, but carries both the amino-  
 CC terminal modification and the internal modification of the Bt ICP  
 CC coding sequence.  
 CC See also AAQ14529, AAQ15142-44.  
 XX SQ Sequence 3201 BP; 915 A; 651 C; 700 G; 930 T; 5 other;  
 Query Match 100.0%; Score 22; DB 12; Length 3201;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcattctacggcgaatgtaccagc 22  
 |||||  
 Db 3021 tcattctacggcgaatgtaccagc 3042  
 RESULT 12  
 AAZ29121  
 ID AAZ29121 standard; DNA; 3336 BP.  
 XX AC AAZ29121;  
 XX DT 21-FEB-2000 (first entry)  
 XX DE Plasmid DVI30 comprising L3/MGFP:NPTII/Tr7 expression cassette.  
 XX KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DVI30;  
 KW GFP:NPTII fusion protein construct; L3/MGFP:NPTII/Tr7; maize R gene;  
 KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
 KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;  
 KW Green fluorescent protein; GFP; gene fusion; selection; screening;  
 KW transformant; expression; automated seed screening technique; assay;  
 KW screenable marker; embryogenic tissue; implementation; ds.  
 XX OS Synthetic.  
 XX PN W09960129-A1.  
 XX PD 25-NOV-1999.  
 XX PF 18-MAY-1999; 99WO-US11023.  
 XX PR 18-MAY-1998; 98US-0080625.  
 XX PF 18-MAY-1999; 99WO-US11023.  
 XX PR 18-MAY-1998; 98US-0080625.  
 XX PA (DEKA-) DEKALB GENETICS CORP.  
 XX PI Kriz AL, Spencer TM;  
 XX DR WPI; 2000-072441/06.  
 XX PT Screenable marker genes useful for identification of transgenic seeds  
 PT for plant breeding -  
 XX PS Example 1; Page 168-170; 182pp; English.  
 XX CC The present DNA sequence is the plasmid DVI33, comprising

PT Screenable marker genes useful for identification of transgenic seeds  
 PT for plant breeding -  
 XX PS Example 1; Page 163-164; 182pp; English.  
 XX CC The present DNA sequence is the plasmid DVI30, that is used in the  
 CC generation of GFP:NPTII fusion protein constructs. It contains an  
 CC expression cassette comprising, a promoter from the maize L3 oleosin  
 CC gene, the coding sequence of MGFP:NPTII translational fusion, excised  
 CC from DVI27 and the Tr7 terminator. This plasmid is used to carry a  
 CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to a gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.  
 XX SQ Sequence 3336 BP; 888 A; 830 C; 764 G; 854 T; 0 other;  
 Query Match 100.0%; Score 22; DB 21; Length 3336;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcattctacggcgaatgtaccagc 22  
 |||||  
 Db 2858 tcattctacggcgaatgtaccagc 2879  
 RESULT 13  
 AAZ29124  
 ID AAZ29124 standard; DNA; 3694 BP.  
 XX AC AAZ29124;  
 XX DT 21-FEB-2000 (first entry)  
 XX DE Plasmid DVI33 used for construction of GFP:NPTII fusion protein.  
 XX KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DVI33;  
 KW GFP:NPTII fusion protein construct; L3/rACT1 intron/EGFP:NPTII/Tr7;  
 KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
 KW EGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;  
 KW Green fluorescent protein; GFP; gene fusion; maize R gene; selection;  
 KW screenable marker; screening; automated seed screening technique; assay;  
 KW transformant; expression; embryogenic tissue; implementation; ds.  
 XX OS Synthetic.  
 XX PN W09960129-A1.  
 XX PD 25-NOV-1999.  
 XX PF 18-MAY-1999; 99WO-US11023.  
 XX PR 18-MAY-1998; 98US-0080625.  
 XX PA (DEKA-) DEKALB GENETICS CORP.  
 XX PI Kriz AL, Spencer TM;  
 XX DR WPI; 2000-072441/06.  
 XX PT Screenable marker genes useful for identification of transgenic seeds  
 PT for plant breeding -  
 XX PS Example 1; Page 168-170; 182pp; English.  
 XX CC The present DNA sequence is the plasmid DVI33, comprising

CC L3/rACT1 intron/EGFP:NPTII/Tr7, that is used in the generation of  
 CC GFP:NPTII fusion protein constructs. It contains an expression cassette  
 CC comprising, a promoter from the maize L3 oleosin gene, the rice actin 1  
 CC intron, the coding sequence of EGFP:NPTII translational fusion, excised  
 CC from DV126 and the Tr7 terminator. This plasmid is used to carry a  
 CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to a gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.

XX  
 SQ Sequence 3694 BP; 890 A; 984 C; 914 G; 906 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3694;  
 Best Local Similarity 100.0%; Pred. No. 0.07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22  
 |||||  
 DB 3216 tcattcagcgcaatgtaccagc 3237

RESULT 14  
 AAZ29123  
 ID AAZ29123 standard; DNA; 3877 BP.  
 XX  
 AC AAZ29123;  
 DT 21-FEB-2000 (first entry)  
 XX  
 DE Plasmid DV132 used for construction of GFP:NPTII fusion protein.  
 XX  
 KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DV132;  
 KW GFP:NPTII fusion protein construct; L3/rACT1 intron/MGFP:NPTII/Tr7;  
 KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
 KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;  
 KW Green fluorescent protein; GFP; gene fusion; maize R gene; selection;  
 KW screenable marker; screening; automated seed screening technique; assay;  
 KW transformant; expression; embryogenic tissue; implementation; ds.  
 XX  
 OS Synthetic.  
 XX WO9960129-A1.  
 XX  
 XX 25-NOV-1999.  
 XX  
 XX 18-MAY-1999; 99WO-US11023.  
 XX  
 XX 18-MAY-1998; 98US-0080625.  
 XX (DEKA-) DEKALB GENETICS CORP.  
 XX Kriz AL, Spencer TM;  
 XX WPI; 2000-072441/06.  
 XX  
 XX Screenable marker genes useful for identification of transgenic seeds  
 XX for plant breeding -  
 XX  
 XX Example 1; Page 166-168; 182pp; English.  
 XX  
 CC The present DNA sequence is the plasmid DV132, comprising  
 CC L3/rACT1 intron/MGFP:NPTII/Tr7, that is used in the generation of  
 CC GFP:NPTII fusion protein constructs. It contains an expression cassette  
 CC comprising, a promoter from the maize L3 oleosin gene, the rice actin 1  
 CC intron, the coding sequence of MGFP:NPTII translational fusion, excised  
 CC from DV127 and the Tr7 terminator. This plasmid is used to carry a

CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to a gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.

XX  
 SQ Sequence 3877 BP; 977 A; 957 C; 922 G; 1021 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3877;  
 Best Local Similarity 100.0%; Pred. No. 0.071;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22  
 |||||  
 DB 3399 tcattcagcgcaatgtaccagc 3420

RESULT 15  
 AAH25423/C  
 ID AAH25423 standard; DNA; 4832 BP.  
 XX  
 AC AAH25423;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 DE Nucleotide sequence of plasmid pTHW118.  
 XX  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barstar gene; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 OS Arabidopsis thaliana.  
 OS Bacillus amyloliquefaciens.  
 OS Nicotiana tabacum.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..25  
 FT /\*tag= a  
 FT /\*note= "right border repeat from TL-DNA from pT1B6S3"  
 FT  
 FT misc\_feature 26..53  
 FT /\*tag= b  
 FT /\*note= "synthetic polylinker derived sequences"  
 FT  
 FT misc\_feature 54..90  
 FT /\*tag= c  
 FT /\*note= "residual sequence from TL-DNA at right  
 FT border repeat"  
 FT  
 FT misc\_feature 91..97  
 FT /\*tag= d  
 FT /\*note= "synthetic polylinker derived sequences"  
 FT  
 FT 3'UTR complement (98..309)  
 FT /\*tag= e  
 FT /\*note= "3' UTR from TL-DNA gene 7 of pT1B6S3"  
 FT  
 FT misc\_feature 310..330  
 FT /\*tag= f  
 FT /\*note= "synthetic polylinker derived sequences"  
 FT  
 FT CDS complement (331..882)  
 FT /\*tag= g  
 FT /\*note= "Streptomyces hygroscopicus bialaphos  
 FT resistance (bar) gene"  
 FT  
 FT promoter complement (883..2608)  
 FT /\*tag= h  
 FT /\*note= "ats1A ribulose-1,5-biphosphate carboxylase  
 FT small subunit gene from Arabidopsis thaliana"  
 FT  
 FT misc\_feature 2609..2658  
 FT /\*tag= i  
 FT /\*note= "synthetic polylinker derived sequences"  
 FT

Search completed: February 15, 2002, 19:01:19  
Job time: 20723 sec

```

FT 3'UTR complement (2659..2919)
FT /*tag= j
FT /note= "Taql fragment from 3' UTR of nopaline
FT synthase gene from T-DNA of pTi37 and
FT containing plant polyadenylation signals"
FT misc_feature 2920..2940
FT /*tag= k
FT /note= "synthetic polylinker derived sequences"
FT 3'UTR 2941..2980
FT /*tag= l
FT /note= "downstream of Bacillus amyloliquefaciens
FT barstar coding region"
FT CDS complement (2981..3253)
FT /*tag= m
FT /note= "Barstar gene coding region from Bacillus
FT amyloliquefaciens"
FT promoter complement (3254..4762)
FT /*tag= n
FT /note= "anther-specific gene TA29 promoter from
FT Nicotiana tabacum"
FT misc_feature 4763..4807
FT /*tag= o
FT /note= "synthetic polylinker derived sequences"
FT misc_feature 4808..4832
FT /*tag= p
FT /note= "left border repeat from TL-DNA from pTiB6S3"
XX
XX WO200141558-A1.
XX
XX 14-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-EP12872.
XX
XX 08-DEC-1999; 99US-0457037.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX De Both G, De Beuckeleer M;
XX WPI; 2001-381419/40.
XX
XX Transgenic winter oilseed rape plants suited for producing hybrid seed
XX with improved qualities, comprises a male-sterility gene and fertility
XX restorer gene, integrated into the genome -
XX
XX Example 1; Page 80-82; 98pp; English.
XX
XX The specification describes a pair of transgenic winter oilseed rape
XX plants suited for producing hybrid seed. One of the plants has an
XX expression cassette comprising a male-sterility gene, and the other
XX plant has an expression cassette comprising a fertility restorer gene,
XX integrated into the genome. The fertility restorer gene is capable of
XX preventing the activity of the male-sterility gene. The plant pair is
XX useful for producing hybrid seed. Plants developed from the hybrid
XX seed have agronomic performance, genetic stability and adaptability to
XX different genetic backgrounds. The present sequence represents
XX plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
XX a fertility restorer gene. The plasmid is used to create transgenic
XX plants of the invention.
XX
XX Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;
XX
Query Match 100.0%; Score 22; DB 22; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcattctacgcgaatgtaccagc 22
DB 247 TCATCTACGGCAATGTACCAGC 226

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:07:32 ; Search time 353.79 Seconds  
(without alignments)  
14.083 Million cell updates/sec

Title: US-09-698-903B-6

Perfect score: 22  
Sequence: 1 tcatctacggaattaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	1186	1	US-08-064-121-2
2	22	100.0	1186	1	US-08-478-015-2
3	22	100.0	1186	3	US-08-475-975-2
4	22	100.0	1186	3	US-09-084-889-2
5	22	100.0	1303	3	US-08-894-440-2
6	22	100.0	3153	4	US-09-080-625-3
7	22	100.0	3200	1	US-08-453-104-23
8	22	100.0	3200	2	US-08-694-824-23
9	22	100.0	3201	1	US-08-453-104-22
10	22	100.0	3201	2	US-08-694-824-22
11	22	100.0	3336	4	US-09-080-625-2
12	22	100.0	3694	4	US-09-080-625-5
13	22	100.0	3877	4	US-09-080-625-4
14	22	100.0	4946	3	US-08-817-188-1
15	22	100.0	5560	3	US-08-817-188-5
16	22	100.0	5864	3	US-08-894-440-4
17	22	100.0	5864	3	US-08-894-440-4
18	22	100.0	6548	3	US-08-894-440-1
19	22	100.0	6548	3	US-08-817-188-2
20	22	100.0	7566	2	US-08-232-016-23
21	22	100.0	7639	2	US-08-232-016-22
22	22	100.0	7811	2	US-08-549-680A-5
23	18.8	85.5	24595	6	5428147-1
24	16	72.7	29	2	US-08-232-016-12
25	15.8	71.8	984	4	US-09-446-504-4
26	15.8	71.8	3574	4	US-09-446-504-83
27	15.8	71.8	3620	4	US-09-446-504-55

c	28	15.2	69.1	238	1	US-07-914-284A-1	Sequence 1, Appli
c	29	15.2	69.1	238	5	PCT-US93-06645-1	Sequence 1, Appli
c	30	14.8	67.3	1290	4	US-09-247-373B-55	Sequence 55, Appl
c	31	14.8	67.3	1347	1	US-08-663-713A-1	Sequence 1, Appli
c	32	14.8	67.3	1347	3	US-09-014-888-1	Sequence 1, Appli
c	33	14.6	66.4	617	2	US-08-392-546C-3	Sequence 3, Appli
	34	14.6	66.4	1512	4	US-08-955-918C-8	Sequence 8, Appli
	35	14.6	66.4	1801	4	US-08-955-918C-6	Sequence 6, Appli
	36	14.6	66.4	1804	1	US-08-631-200-1	Sequence 1, Appli
	37	14.6	66.4	1804	1	US-08-829-553-1	Sequence 1, Appli
	38	14.6	66.4	1804	2	US-08-922-267A-1	Sequence 1, Appli
	39	14.6	66.4	1804	2	US-08-936-707A-1	Sequence 1, Appli
	40	14.6	66.4	1804	2	US-08-936-706A-1	Sequence 1, Appli
	41	14.6	66.4	1804	3	US-09-248-203-1	Sequence 1, Appli
	42	14.6	66.4	1804	4	US-09-406-071-1	Sequence 1, Appli
	43	14.6	66.4	2119	1	US-08-630-592-1	Sequence 1, Appli
	44	14.6	66.4	2119	1	US-08-714-991-1	Sequence 1, Appli
	45	14.6	66.4	2119	3	US-09-032-365A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-064-121-2  
; Sequence 2, Application us/08064121  
; Patent No. 5641664  
; GENERAL INFORMATION:  
; APPLICANT: D'HALLUIN, Kathleen  
; APPLICANT: Gobel, Elke  
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/064,121  
; FILING DATE: 24-MAY-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 90403332.1  
; FILING DATE: 23-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91401888.2  
; FILING DATE: 08-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crane-Feury, Sharon E  
; REGISTRATION NUMBER: 36,113  
; REFERENCE/DOCKET NUMBER: 010830-043  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1186 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: probe

```

;
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
;
US-08-064-121-2

Query Match 100.0%; Score 22; DB 1; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 2
US-08-478-015-2
; Sequence 2, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401898.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:

```

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine phosphotransferase ge
;
US-08-478-015-2

```

```

Query Match 100.0%; Score 22; DB 1; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

```

```

RESULT 3
US-08-475-975-2
; Sequence 2, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sta.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:

```

```
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
;
US-08-475-975-2

Query Match 100.0%; Score 22; DB 3; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 4
US-09-889-2
; Sequence 2, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: Gobel, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/064,121
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
;
US-09-084-889-2

Query Match 100.0%; Score 22; DB 3; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 5
US-08-894-440-2
; Sequence 2, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
; OTHER INFORMATION: fragment of pTS88
```

```

RESULT      6
US-09-080-625-3
; Sequence 3, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; APPLICANT: Spencer, T. Michael
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

Query Match 100.0%; Score 22; DB 1; Length 3200;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
|||||

Db 2948 TCATCTACGGCAATGTACCAGC 2969

RESULT 8  
US-08-694-824-23  
; Sequence 23, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 100.0%; Score 22; DB 2; Length 3200;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
|||||

Db 2948 TCATCTACGGCAATGTACCAGC 2969

RESULT 9  
US-08-453-104-22  
; Sequence 22, Application US/08453104

; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 100.0%; Score 22; DB 1; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
|||||

Db 3021 TCATCTACGGCAATGTACCAGC 3042

RESULT 10  
US-08-694-824-22  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

```
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,824
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2151..2155
; OTHER INFORMATION: /note= "Nucleotides 2151-2155
; OTHER INFORMATION: wherein N is not known."
;
US-08-694-824-22

Query Match 100.0%; Score 22; DB 2; Length 3201;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcatctacggcaatgtaccagc 22
|||||
Db 3021 TCATCTACGGCAATGTACCAGC 3042

RESULT 11
US-09-080-625-2
; Sequence 2, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-09-080-625-5
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3336 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-09-080-625-2

Query Match 100.0%; Score 22; DB 4; Length 3336;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcatctacggcaatgtaccagc 22
|||||
Db 2858 TCATCTACGGCAATGTACCAGC 2879

RESULT 12
US-09-080-625-5
; Sequence 5, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; APPLICANT: Spencer, T. Michael
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-09-080-625-5
```



Query Match 100.0%; Score 22; DB 4; Length 3694;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattacggcaatgtaccagc 22  
|||||  
Db 3216 TCATCTACGGCAATGTACCAGC 3237

## RESULT 13

US-09-080-625-4  
; Sequence 4, Application US/09080625  
; Patent No. 6307123  
; GENERAL INFORMATION:  
; APPLICANT: Kriz, Alan L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
; TITLE OF INVENTION: IDENTIFICATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/080,625  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, Robert E.  
; REGISTRATION NUMBER: P-42,628  
; REFERENCE/DOCKET NUMBER: DEKM:161  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3877 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-080-625-4

Query Match 100.0%; Score 22; DB 4; Length 3877;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattacggcaatgtaccagc 22  
|||||  
Db 3399 TCATCTACGGCAATGTACCAGC 3420

## RESULT 14

US-08-817-188-1/c  
; Sequence 1, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; CURRENT FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
; EARLIER FILING DATE: 1996-07-31

; EARLIER APPLICATION NUMBER: EP 95401844.6  
; EARLIER FILING DATE: 1995-08-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4946  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; OTHER INFORMATION: plasmid pTHW107  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((1)..(25))  
; OTHER INFORMATION: T-DNA right border (RB)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((97)..(330))  
; OTHER INFORMATION: 3' g7; 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((331)..(882))  
; OTHER INFORMATION: bar; region coding for phosphinotricin acetyl  
; OTHER INFORMATION: transferase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((883)..(2608))  
; OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2658)..(3031))  
; OTHER INFORMATION: 3' nos; 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
; OTHER INFORMATION: gene of Agrobacterium T-DNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((3032)..(3367))  
; OTHER INFORMATION: barnase; region coding for barnase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((3368)..(4876))  
; OTHER INFORMATION: PTA29; promoter region of TA29 gene of Nicotiana  
; OTHER INFORMATION: tabacum  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((4922)..(4946))  
; OTHER INFORMATION: LB; T-DNA left border  
US-08-817-188-1

Query Match 100.0%; Score 22; DB 3; Length 4946;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattacggcaatgtaccagc 22  
|||||  
Db 247 TCATCTACGGCAATGTACCAGC 226

## RESULT 15

US-08-817-188-5/c  
; Sequence 5, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; CURRENT FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366

EARLIER FILING DATE: 1996-07-31  
EARLIER APPLICATION NUMBER: EP 95401844.6  
EARLIER FILING DATE: 1995-08-04  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 5560  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
OTHER INFORMATION: plasmid pTHW142  
NAME/KEY: misc\_feature  
LOCATION: (1)..(25)  
OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from  
OTHER INFORMATION: pTiB6S3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (84)..(296)  
OTHER INFORMATION: 3', 97: 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
OTHER INFORMATION: T-DNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (318)..(869)  
OTHER INFORMATION: bar: region coding for phosphinotricin  
OTHER INFORMATION: acetyltransferase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (830)..(2760)  
OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit  
OTHER INFORMATION: gene of Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2765)..(3058)  
OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript  
OTHER INFORMATION: containing polyadenylation signals  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3059)..(5056)  
OTHER INFORMATION: uidA: region coding for beta-glucuronidase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (4483)..(4671)  
OTHER INFORMATION: IV2: region corresponding to the second intron of  
OTHER INFORMATION: the ST-LS1 gene  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (5067)..(5502)  
OTHER INFORMATION: P35S: 35S promoter region of CamV  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (5533)..(5560)  
OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from  
OTHER INFORMATION: pTiB6S3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (5058)..(5059)  
OTHER INFORMATION: region with unknown sequence (may contain up to 20  
OTHER INFORMATION: nucleotides)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (5077)..(5078)  
OTHER INFORMATION: region with unknown sequence (may contain up to 20  
OTHER INFORMATION: nucleotides)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (5476)..(5479)  
OTHER INFORMATION: region with unknown sequence (may contain up to 20  
OTHER INFORMATION: nucleotides)

US-08-817-188-5

Query Match 100.0%; Score 22; DB 3; Length 5560;  
Best Local Similarity 100.0%; Pred. No. 0.026;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
Qy 1 tcattctacggcaatgtaccagc 22  
|||||  
Db 234 TCATCTAGGCAATGTACCAGC 213

Search completed: February 15, 2002, 19:07:35  
Job time: 14629 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:03:51 ; Search time 9904.61 Seconds  
(without alignments)  
23.868 Million cell updates/sec

Title: US-09-698-903B-6  
Perfect score: 22  
Sequence: 1 tcatctacgcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_estl:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	17.4	79.1	568	11	B1174209
2	17.2	78.2	330	10	AW358852
3	17.2	78.2	352	10	AA475114
4	17.2	78.2	355	10	BE032643
5	17.2	78.2	355	11	BE924637
6	17.2	78.2	432	10	AA212655
7	17.2	78.2	562	13	AQ724420
8	17.2	78.2	567	13	AZ851447
9	17.2	78.2	721	11	BG592665
10	17.2	78.2	721	11	BG964419
c 11	17.2	78.2	932	11	BG400800
12	16.8	76.4	162	13	AZ474239

13	16.8	76.4	855	11	BG669143
c 14	16.4	74.5	479	10	AA458389
c 15	16.2	73.6	216	10	BE168752
c 16	16.2	73.6	309	13	TA105D12Q
17	16.2	73.6	314	13	AZ319004
18	16.2	73.6	343	10	AW789094
c 19	16.2	73.6	352	11	T69520
20	16.2	73.6	413	10	AA637160
21	16.2	73.6	423	10	AI593113
22	16.2	73.6	428	10	AI280109
c 23	16.2	73.6	435	11	T88735
24	16.2	73.6	438	10	AV668989
25	16.2	73.6	440	11	N62954
26	16.2	73.6	443	10	AA878226
c 27	16.2	73.6	450	10	AA256224
28	16.2	73.6	450	10	AA574351
c 29	16.2	73.6	457	11	BG730070
c 30	16.2	73.6	495	10	AI449899
31	16.2	73.6	501	10	AI338197
c 32	16.2	73.6	502	11	BI234655
33	16.2	73.6	505	13	AZ720352
c 34	16.2	73.6	508	10	AI862985
c 35	16.2	73.6	544	10	AI533411
36	16.2	73.6	553	10	AI580125
c 37	16.2	73.6	571	10	BE367030
c 38	16.2	73.6	609	10	BE367017
c 39	16.2	73.6	611	11	W37106
c 40	16.2	73.6	622	11	BE819452
c 41	16.2	73.6	644	11	BG049365
c 42	16.2	73.6	658	10	AW953929
43	16.2	73.6	701	10	BE542678
44	16.2	73.6	739	11	BF177958
45	16.2	73.6	753	10	BE393004

ALIGNMENTS

RESULT 1  
B1174209/c  
LOCUS B1174209 568 bp mRNA EST 09-JUL-2001  
DEFINITION OSTF013F8\_1 AD-wrmcDNA Caenorhabditis elegans cDNA similar to C46F11.2, mRNA sequence.  
ACCESSION B1174209  
VERSION B1174209.1 GI:14640012  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea ; Rhabditiidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 568)  
AUTHORS Reboul,J., Vaglio,P., Tzellas,N., Thierry-Mieg,N., Moore,T., Jackson,C., Shin-I,T., Kohara,Y., Thierry-Mieg,D., Thierry-Mieg,J., Lee,H., Hitt,J., Doucette-Stamm,L., Hartley,J.L., Temple,G.F., Brasch,M.A., Vandenhaute,J., Lamesch,P.E., Hill,D.E. and Vidal,M.  
TITLE Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in C. elegans  
JOURNAL Nat. Genet. 27 (3), 332-336 (2001)  
MEDLINE 21135099  
COMMENT Contact: Reboul J, Vaglio P  
Marc Vidal Laboratory  
Dana Farber Cancer Institute  
44 Binney Street, Boston, MA 02115, USA  
Tel: 617 632 5180  
Fax: 617 632 2425  
Email: Jerome\_Reboul@dfci.harvard.edu  
Sequence tag of Gateway entry clones. The primers used were designed on the predicted protein encoding ORF. C. elegans ORFome cloning project : Contact jerome\_reboul@dfci.harvard.edu or philippe\_vaglio@dfci.harvard.edu  
POLYA-No. Location/Qualifiers  
1. .568 source





polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGAGCGCGCGAATCTTTTTTTTTTTT 3'] (Pharmacia), digested with Not I and Eco RI adaptors and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Patima Bonaldo.\*

BASE COUNT 71 a 148 c 125 g 88 t  
ORIGIN

Query Match 78.2%; Score 17.2; DB 10; Length 432;  
Best Local Similarity 86.4%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattcaggcaattaccagc 22  
||||||| | |||||

Db 94 TCATCTACCGGATTGACCAGC 115  
||||||| | |||||

RESULT 7  
A0724420 562 bp DNA GSS 14-JUL-1999  
LOCUS HS\_2119\_A2\_C07\_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2119 Col=14 Row=E, DNA sequence.  
DEFINITION A0724420  
ACCESSION A0724420  
VERSION A0724420.1 GI:5484089  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 562)  
Mahairas,G.G., Wallace,J.C., Smith,K., Svartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc.Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3867  
Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 2119 row: E column: 14  
Seq primer: M13 Reverse  
Class: BAC ends

High quality sequence stop: 562.  
Location/Qualifiers

1..562  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=2119 Col=14 Row=E"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 142 a 137 c 122 g 153 t 8 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 13; Length 562;  
Best Local Similarity 86.4%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattcaggcaattaccagc 22  
||||||| | |||||

Db 438 TCATCTAGCCACGTGTGCCAGC 459

RESULT 8  
A2851447/c

LOCUS A2851447.1

DEFINITION A2851447.1

ACCESSION A2851447

VERSION A2851447.1

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS 1 (bases 1 to 567)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0153 row: M column: 11  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 567.  
Location/Qualifiers

1..567  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0153M11"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 188 a 94 c 106 g 179 t  
ORIGIN

Query Match 78.2%; Score 17.2; DB 13; Length 567;  
Best Local Similarity 86.4%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattcaggcaattaccagc 22



```

Db 300 TCAACTACAGCAATGAACAGC 279
||||| ||||| ||||| |||||
RESULT 9
BG592665 721 bp mRNA EST 12-APR-2001
LOCUS EST491343 cSTS Solanum tuberosum cDNA clone cSTS2C3 5' sequence,
DEFINITION mRNA sequence.
ACCESSION BG592665
VERSION BG592665.1 GI:13610805
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 721)
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,
Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE Generations of ESTs from sprouting potato eyes
JOURNAL Unpublished (2000)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13P-R.
FEATURES
source
1..721
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTS2C3"
/clone_lib="cSTS"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/notes="vector: pbluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."
BASE COUNT 192 a 202 c 113 g 214 t
ORIGIN
Query Match 78.2%; Score 17.2; DB 11; Length 721;
Best Local Similarity 86.4%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 tcattacggcaatgtaccagc 22
||||| ||||| ||||| |||||
Db 282 TCGTCCACCGCAATGTACCAGC 303
||||| ||||| ||||| |||||
RESULT 10
BG964419 721 bp mRNA EST 12-JUN-2001
LOCUS BG964419
DEFINITION 602832063F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4986807 5',
mRNA sequence.
ACCESSION BG964419
VERSION BG964419.1 GI:14352056
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 721)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

```

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M10996 row: m column: 16
High quality sequence stop: 718.
FEATURES
source
1..721
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4986807"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 124 a 223 c 228 g 146 t
ORIGIN
Query Match 78.2%; Score 17.2; DB 11; Length 721;
Best Local Similarity 86.4%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 tcattacggcaatgtaccagc 22
||||| ||||| ||||| |||||
Db 611 TCATCTACCGGATGTACCAGC 632
||||| ||||| ||||| |||||
RESULT 11
BG400800 932 bp mRNA EST 12-MAR-2001
LOCUS BG400800
DEFINITION 602464016F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592474 5',
mRNA sequence.
ACCESSION BG400800
VERSION BG400800.1 GI:13294248
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 932)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1330 row: o column: 03
High quality sequence stop: 230.
FEATURES
source
1..932
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4592474"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccatgtgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGCCC-3' and 3' adaptor sequence:

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0290J11"
/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative

```

.Qy 3 atctacggcaatgtaccagc 22  
||||| ||||| |||||  
Db 673 ATCTAGGCAATGTCCAGC 692

RESULT 14  
AA458389/c  
LOCUS  
DEFINITION  
v49f10.t1 Soares\_mammary\_gland\_NbMMG Mus musculus cDNA clone  
IMAGE:864715 5' similar to gb:J03040 SPARC PRECURSOR (HUMAN);  
qb:M20692 Mouse osteonectin (MOUSE);, mRNA sequence.  
ACCESSION  
AA458389  
VERSION  
AA458389.1 GI:2181109  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE  
1 (bases 1 to 479)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE  
The WashU-HMMI Mouse EST Project  
JOURNAL  
Unpublished (1996)  
COMMENT  
Contact: Marra M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL : contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:508803  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 441.

FEATURES  
source  
1..479  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:864715"  
/clone\_lib="Soares\_mammary\_gland\_NbMMG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5'  
TGTTACCAATCTGAAGTGGGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Ronaldto."  
BASE COUNT 114 a 125 c 127 g 112 t 1 others  
ORIGIN

Query Match 74.5%; Score 16.4; DB 10; Length 479;  
Best Local Similarity 94.4%; Pred. No. 1.1e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 tcattacggcaatgtac 18  
||||| ||||| |||||  
Db 312 TCATCCAGCGCAATGTAC 295

RESULT 15  
BE168752/c  
LOCUS  
DEFINITION  
OVI-HT0516-140300-107-e07 HT0516 Homo sapiens cDNA, mRNA sequence.  
ACCESSION  
BE168752  
VERSION  
BE168752.1 GI:8631473  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 216)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.J.  
TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE  
20202663  
COMMENT  
Contact: Simpson A.J.G  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV1-HT0516-140  
300-107-e07&t3=2000-03-14&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 216.  
FEATURES  
source  
1..216  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0516"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 42 a 75 c 59 g 40 t  
ORIGIN  
Query Match 73.6%; Score 16.2; DB 10; Length 216;  
Best Local Similarity 85.7%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 tcattacggcaatgtaccag 21  
||||| ||||| |||||  
Db 46 TCAGCCAGCGCAAGTACCAG 26  
Search completed: February 15, 2002, 18:03:55  
Job time: 20944 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:47:39 ; Search time 2553.1 Seconds  
(without alignments)  
155.079 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24  
Sequence: 1 tcagaagtatcagcagctccacc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_cm.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pi.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	24	6	AX127756
2	24	100.0	270	6	A71431 Sequence 1
3	24	100.0	323	6	A7282 Sequence 7
4	24	100.0	340	6	A21284 Artificial
5	24	100.0	474	1	BABARSTA
6	24	100.0	4032	6	A71435 Sequence 5
7	24	100.0	4808	6	AR007527 Sequence
8	24	100.0	4808	6	AR084093 Sequence
9	24	100.0	4832	6	AX172441 Sequence
10	24	100.0	5865	6	AX127748 Sequence
11	24	100.0	6555	6	AR007512 Sequence
12	24	100.0	6555	6	AR084078 Sequence
13	20.8	86.7	273	6	A71433 Sequence 3
14	20.8	86.7	563	6	A71436 Sequence 6
15	20.8	86.7	5349	6	A71437 Sequence 7
16	20.8	86.7	5611	6	A71440 Sequence 10
17	19.8	82.5	130235	8	AC008007 Genomic s
18	18.2	75.8	426	9	AF062270 Homo sapi
19	18.2	75.8	74213	2	AC025320 Homo sapi
20	18.2	75.8	124990	9	HS434P1
21	18.2	75.8	165471	2	AC015713 Homo sapi
22	18.2	75.8	234131	2	AC093357 Mus muscu
23	17.8	74.2	1905	1	COXCYTSYN M36338 C.burnetii
24	17.8	74.2	173846	2	AC087781 Mus muscu
25	17.8	74.2	238737	2	AC084073 Mus muscu
26	17.6	73.3	731	9	HSDCTN08
27	17.6	73.3	3534	6	AX015400 Sequence
28	17.6	73.3	4153	9	HSDYNACTN
29	17.6	73.3	12426	9	HSDCTN1A3
30	17.6	73.3	12956	1	AE006198 Pasteurel
31	17.6	73.3	13223	1	AEVU49859
32	17.6	73.3	34646	3	U00066 Caenorhabdi
33	17.6	73.3	46843	8	SPB530
34	17.6	73.3	50502	9	HSAA94016
35	17.6	73.3	52358	9	HS27C10
36	17.6	73.3	69506	8	AC012680 Arabidops
37	17.6	73.3	95769	8	AC013430 Genomic s
38	17.6	73.3	129538	9	AC008854 Homo sapi
39	17.6	73.3	139152	8	AP002525 Oryza sat
40	17.6	73.3	141198	2	AC023898 Mus muscu
41	17.6	73.3	147722	2	AC016783 Homo sapi
42	17.6	73.3	157993	2	AC073354 Homo sapi
43	17.6	73.3	167263	2	AC063966 Homo sapi
44	17.6	73.3	167943	2	AC026322 Homo sapi
45	17.6	73.3	170081	2	AC055714 Homo sapi

ALIGNMENTS

RESULT 1					
AX127756	AX127756	24 bp	DNA	PAT	15-MAY-2001
LOCUS	Sequence 9 from Patent WO0131042.				
DEFINITION	AX127756				
ACCESSION	AX127756.1	GI:14134403			
VERSION					
KEYWORDS	synthetic construct.				
SOURCE	synthetic construct				
ORGANISM	artificial sequence.				
REFERENCE	1 (bases 1 to 24)				
AUTHORS	Weston,B. and de Beuckeleer,M.				
TITLE	Male-sterile brassica plants and methods for producing same				
JOURNAL	Patent: WO 0131042-A 9 03-MAY-2001;				
FEATURES	Aventis CropScience N.V. (BE)				
source	Location/Qualifiers				
	1..24				
	/organism="synthetic construct"				
	/db_xref="taxon:32630"				
	/note="primer MDB8"				

BASE COUNT 7 a 9 c 4 g 4 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 24;  
Best Local Similarity 100.0%; Pred. No. 0.086;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 1 TCAGAAAGTATCAGCGACCTCCACC 24

## RESULT 2

LOCUS A71431 270 bp DNA PAT 07-MAY-1999  
DEFINITION Sequence 1 from Patent WO9810081.  
ACCESSION A71431  
VERSION A71431.1 GI:4775044

KEYWORDS  
SOURCE Bacillus amyloliquefaciens.  
ORGANISM Bacillus amyloliquefaciens

REFERENCE 1 (bases 1 to 270)  
Bacillus/Staphylococcus group; Bacillus.

AUTHORS Michiels, F. and Williams, M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 1 12-MAR-1998;  
MICHELIS FRANK (BE)

FEATURES  
source Location/Qualifiers  
1..270  
/organism="Bacillus amyloliquefaciens"  
/db\_xref="taxon:1390"

CDS  
1..270  
/function="INHIBITOR OF BARNASE"  
/codon\_start=1  
/transl\_table=11  
/product="BARSTAR"  
/protein\_id="CAB42577.1"  
/db\_xref="GI:4775045"  
/translation="MKKAVINGEQIRISDLHQTLLKELALPEYYGENLDALWDCLTG  
WVEYPLVLEWRFQESKQLTENGAEVSLQVFEAKAGCDITILS"

BASE COUNT 80 a 58 c 74 g 58 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 270;  
Best Local Similarity 100.0%; Pred. No. 0.095;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 32 TCAGAAAGTATCAGCGACCTCCACC 55

## RESULT 3

LOCUS A87282 323 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 7 from Patent WO9837211.  
ACCESSION A87282  
VERSION A87282.1 GI:6736047

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 323)  
Huttner, E. and Betzner, A.S.  
AUTHORS PROTEIN COMPLEMENTATION IN TRANSGENIC PLANTS  
TITLE Patent: WO 9837211-A 7 27-AUG-1998;  
JOURNAL GENE SHEARS PTY LTD (AU); HUTTNER ERIC (AU)

FEATURES  
source Location/Qualifiers  
1..323  
/organism="unidentified"

## CDS

/db\_xref="taxon:32644"  
1..273  
/codon\_start=1  
/product="BARSTAR"  
/protein\_id="CAB69371.1"  
/db\_xref="GI:6736048"  
/translation="MKKAVINGEQIRISDLHQTLLKELALPEYYGENLDALWDCLTG  
WVEYPLVLEWRFQESKQLTENGAEVSLQVFEAKAGCDITILS"

BASE COUNT 98 a 69 c 87 g 69 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 323;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 32 TCAGAAAGTATCAGCGACCTCCACC 55

## RESULT 4

LOCUS A21284 340 bp DNA PAT 31-MAY-1994  
DEFINITION Artificial barstar gene.  
ACCESSION A21284  
VERSION A21284.1 GI:514151

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified

REFERENCE 1 (bases 1 to 340)  
Marians, C., Leemans, J. and De Greef, W.  
AUTHORS Plants with modified flowers  
TITLE Patent: EP 0412911-A 2 13-FEB-1991;  
JOURNAL PLANT GENETIC SYSTEMS, N.V.

FEATURES  
source Location/Qualifiers  
1..340  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 106 a 73 c 92 g 69 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 340;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 42 TCAGAAAGTATCAGCGACCTCCACC 65

## RESULT 5

LOCUS BABARSTA 474 bp DNA BCT 23-JUN-1996  
DEFINITION Bacillus amyloliquefaciens barstar gene.  
ACCESSION X15545  
VERSION X15545.1 GI:1155006

KEYWORDS barstar; ribonuclease inhibitor.  
SOURCE Bacillus amyloliquefaciens.

ORGANISM Bacillus/Staphylococcus group; Bacillus.  
REFERENCE 1 (bases 1 to 436)  
Bacillus/Staphylococcus group; Bacillus.

AUTHORS Hartley, R.W.  
TITLE Barnase and barstar. Expression of its cloned inhibitor permits  
expression of a cloned ribonuclease  
J. Mol. Biol. 202 (4), 913-915 (1988)  
JOURNAL 89012012  
MEDLINE  
REFERENCE 2 (bases 1 to 474)  
Hartley, R.  
AUTHORS Direct Submission  
TITLE

JOURNAL Submitted (14-JAN-1996) R.Hartley, LCDB/NIDDK, NIH, Bethesda, 20892  
USA, email:hartley@helix.nih.gov  
REMARK Revised by author  
COMMENT On Jan 15, 1996 this sequence version replaced gi:39311.  
See also acc# x12871.  
FEATURES  
source 1..474  
Location/Qualifiers  
/organism="Bacillus amyloliquefaciens"  
/db\_xref="taxon:1390"  
/clone="pMT311"  
94..99  
/note="pot. ribosome binding site"  
109..119  
124..396  
/note="barstar (AA 1 - 90)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA33551.1"  
/db\_xref="GI:39312"  
/db\_xref="SWISS-PROT:P11540"  
/translation="MKKAVINGEQIRISLDLHQLKKELALPEYIGENLDALWDCLTG  
WVEYPIVLWEQFQSKLTENGAEVLOVFRKAEGCDITILS"  
BASE COUNT 154 a 104 c 123 g 93 t  
ORIGIN

Query Match 100.0%; Score 24; DB 1; Length 474;  
Best Local Similarity 100.0%; Pred. No. 0.098;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 155 TCAGAAGTATCAGCGACCTCCACC 178

RESULT 6  
A71435  
LOCUS A71435 4032 bp DNA circular PAT 07-MAY-1999  
DEFINITION Sequence 5 from Patent WO9810081.  
ACCESSION A71435  
VERSION A71435.1 GI:4775048  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified.  
REFERENCE  
AUTHORS Michiels,F. and Williams,M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 5 12-MAR-1998;  
MICHIELS FRANK (BE)  
FEATURES  
source 1..4032  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 1072 a 968 c 963 g 1029 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4032;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 3435 TCAGAAGTATCAGCGACCTCCACC 3458

RESULT 7  
AR007527/c  
LOCUS AR007527 4808 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 17 from patent US 5750867.  
ACCESSION AR007527

VERSION AR007527.1 GI:3967011  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 4808)  
AUTHORS Williams,M. and Leemans,J.  
TITLE Maintenance of male-sterile plants  
JOURNAL Patent: US 5750867-A 17 12-MAY-1998;  
FEATURES Location/Qualifiers  
1..4808  
source /organism="unknown"  
BASE COUNT 1370 a 1063 c 1038 g 1333 t 4 others  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4808;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 8  
AR084093/c  
LOCUS AR084093 4808 bp DNA PAT 01-SEP-2000  
DEFINITION Sequence 17 from patent US 5977433.  
ACCESSION AR084093  
VERSION AR084093.1 GI:10010864  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 4808)  
AUTHORS Williams,M. and Leemans,J.  
TITLE Maintenance of male-sterile plants  
JOURNAL Patent: US 5977433-A 17 02-NOV-1999;  
FEATURES Location/Qualifiers  
1..4808  
source /organism="unknown"  
BASE COUNT 1370 a 1063 c 1038 g 1333 t 4 others  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4808;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 9  
AX172441/c  
LOCUS AX172441 4832 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 2 from Patent WO0141558.  
ACCESSION AX172441  
VERSION AX172441.1 GI:14597553  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 4832)  
AUTHORS de Both,G. and de Beuckeleer,M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 2 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES Location/Qualifiers  
1..4832  
source /organism="synthetic construct"

misc\_feature  
BASE COUNT 1528 a 883 c 932 g 1488 t 1 others  
ORIGIN  
/db\_xref="taxon:32630"  
/note="T-DNA of plasmid pTHW118"  
1883...4065  
/note="HpaI restriction fragment"

Query Match  
Best Local Similarity 100.0%; Score 24; DB 6; Length 4832;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 3222 TCAGAAATATCAGCGACCTCCACC 3199

RESULT 10  
AX127748  
LOCUS AX127748 5865 bp DNA PAT 15-MAY-2001  
DEFINITION Sequence 1 from Patent WO0131042.  
ACCESSION AX127748  
VERSION AX127748.1 GI:14134395  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 5865)  
AUTHORS Weston, B. and de Heuckeleer, M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 1 03-MAY-2001;  
FEATURES  
source  
1. 5865  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="T-DNA of plasmid pCO113"  
BASE COUNT 1849 a 1095 c 1149 g 1772 t  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 24; DB 6; Length 5865;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 5249 TCAGAAATATCAGCGACCTCCACC 5272

RESULT 11  
AR007512/c  
LOCUS AR007512 6555 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 2 from patent US 5750867.  
ACCESSION AR007512  
VERSION AR007512.1 GI:3966996  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 6555)  
AUTHORS Williams, M. and Leemans, J.  
TITLE Maintenance of male-sterile plants  
JOURNAL Patent: US 5750867-A 2 12-MAY-1998;  
FEATURES  
source  
1. 6555  
Location/Qualifiers  
/organism="unknown"  
BASE COUNT 1690 a 1611 c 1584 g 1670 t  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 24; DB 6; Length 6555;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 993 TCAGAAATATCAGCGACCTCCACC 970

RESULT 12  
AR084078/c  
LOCUS AR084078 6555 bp DNA PAT 01-SEP-2000  
DEFINITION Sequence 2 from patent US 5977433.  
ACCESSION AR084078  
VERSION AR084078.1 GI:10010849  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 6555)  
AUTHORS Williams, M. and Leemans, J.  
TITLE Maintenance of male-sterile plants  
JOURNAL Patent: US 5977433-A 2 02-NOV-1999;  
FEATURES  
source  
1. 6555  
Location/Qualifiers  
/organism="unknown"  
BASE COUNT 1690 a 1611 c 1584 g 1670 t  
ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 24; DB 6; Length 6555;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 993 TCAGAAATATCAGCGACCTCCACC 970

RESULT 13  
A71433  
LOCUS A71433 273 bp DNA PAT 07-MAY-1999  
DEFINITION Sequence 3 from Patent WO9810081.  
ACCESSION A71433  
VERSION A71433.1 GI:4775046  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 273)  
AUTHORS Michiels, F. and Williams, M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 3 12-MAR-1998;  
FEATURES  
source  
1. 273  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"  
1. >273  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAB42578.1"  
/db\_xref="GI:4775047"

BASE COUNT 62 a 84 c 85 g 42 t  
ORIGIN  
/translation="MAKRAVINGEIRISDSIHQTIKKELALPEYYGENDALWDCLT  
GWVEYPLVLEWRFQSQKLTENGAEVQLQVFEAKAGCDITILS"

Query Match  
Best Local Similarity 91.7%; Score 20.8; DB 6; Length 273;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 tcagaagtatcagcgacctccacc 24  
|||||



Db 35 TCAGGAGCATCAGCGACCTCCACC 58

RESULT 14

A71436 A71436 563 bp DNA PAT 07-MAY-1999  
LOCUS Sequence 6 from Patent WO9810081.  
DEFINITION A71436  
ACCESSION A71436  
VERSION A71436.1 GI:4775049  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 563)  
AUTHORS Michiels,F. and Williams,M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 6 12-MAR-1998;  
MICHIELS FRANK (BE)  
FEATURES Location/Qualifiers  
source 1..563  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 152 a 133 c 145 g 133 t  
ORIGIN

Query Match 86.7%; Score 20.8; DB 6; Length 563;  
Best Local Similarity 91.7%; Pred. No. 4.8;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24

||||| | | | | | | | | | | | | | | |

Db 37 TCAGGAGCATCAGCGACCTCCACC 60

RESULT 15

A71437 A71437 5349 bp DNA PAT 07-MAY-1999  
LOCUS Sequence 7 from Patent WO9810081.  
DEFINITION A71437  
ACCESSION A71437  
VERSION A71437.1 GI:4775050  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 5349)  
AUTHORS Michiels,F. and Williams,M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 7 12-MAR-1998;  
MICHIELS FRANK (BE)  
FEATURES Location/Qualifiers  
source 1..5349  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 1339 a 1233 c 1290 g 1487 t  
ORIGIN

Query Match 86.7%; Score 20.8; DB 6; Length 5349;  
Best Local Similarity 91.7%; Pred. No. 5.3;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24

||||| | | | | | | | | | | | | | | |

Db 4004 TCAGGAGCATCAGCGACCTCCACC 4027

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:01:52 ; Search time 868.33 Seconds  
(without alignments)  
23.696 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24

Sequence: 1 tcagaagtatcagcgacctccacc 24

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	22	PCR primer MDB251
2	24	100.0	270	19	Wild type barstar
3	24	100.0	323	19	Barstar coding seq
4	24	100.0	340	12	Clai-HindIII fragme
5	24	100.0	1303	17	Plasmid pTS88 (Eco
6	24	100.0	2275	22	Oligonucleotide #1
7	24	100.0	3544	17	PTS200 contg. p35S
8	24	100.0	3544	20	Nucleotide sequenc
9	24	100.0	4032	19	Plasmid pMW71. Sy
10	24	100.0	4808	15	Restriction fragme
11	24	100.0	4832	22	Nucleotide sequenc

12	24	100.0	4896	17	AAT08976	PTS256 contg. p35S
13	24	100.0	4896	20	AA15631	Nucleotide sequenc
14	24	100.0	5864	17	AAT39339	Plasmid pTCO113 T-
15	24	100.0	5865	22	AAD06990	Chimeric T-DNA of
16	24	100.0	5555	15	AA053874	Plasmid pVE144 use
17	24	100.0	7432	22	AAE86441	Plasmid pTS346. U
18	20.8	86.7	273	19	AAV23235	Synthetic barstar
19	20.8	86.7	563	19	AAV23238	Part of plasmid pL
20	20.8	86.7	5349	19	AAV23239	T-DNA of pPTS24.
21	20.8	86.7	5611	19	AAV23242	Plasmid pLH48. Sy
22	17.6	73.3	245	22	AA152182	Probe #20868 used
23	17.6	73.3	469	22	AA139127	Probe #7813 used t
24	17.6	73.3	3534	20	AA241289	Human normal ovari
25	17.6	73.3	4721	22	AA159601	Human polynucleoti
26	17.6	73.3	4753	22	AA157815	Human polynucleoti
27	17.2	71.7	384	22	AA120952	Probe #10885 for g
28	17.2	71.7	384	22	AA146196	Probe #14882 used
29	17.2	71.7	384	22	AA106663	Probe #6654 used t
30	17.2	71.7	450	22	AA111739	Probe #1672 for ge
31	17.2	71.7	450	22	AA133046	Probe #1732 used t
32	17.2	71.7	450	22	AA101667	Probe #1658 used t
33	17.2	71.7	543	21	AAC34918	Arabidopsis thalia
34	17.2	71.7	616	21	AAC50006	Arabidopsis thalia
35	17.2	71.7	10223	19	AAV52206	Streptococcus pneu
36	17.2	71.7	160552	22	AAD02697	Human glycosyl sul
37	17	70.8	1842	22	AAC83227	DNA sequence from
38	16.8	70.0	3831	19	AAV52424	Streptococcus pneu
39	16.6	69.2	154	21	AA889494	Cone snail alpha-c
40	16.6	69.2	609	22	AAH67560	C glutamicum codin
41	16.6	69.2	1239	22	AAH45076	Rat FM-3 coding se
42	16.6	69.2	1292	22	AAD08025	Rat G-protein coup
43	16.6	69.2	1371	22	AAD08026	Rat G-protein coup
44	16.6	69.2	1650	22	AAE33121	Human secreted pro
45	16.6	69.2	1866	17	AAH41492	DNA encoding MAGE-

#### ALIGNMENTS

RESULT 1

AAD06998  
ID AAD06998 standard; DNA; 24 BP.

AC AAD06998;

XX XX

DT 06-AUG-2001 (first entry)

XX XX

DE PCR primer MDB251 to generate the flanking region of elite event MS-B2.

XX MS-B2 elite event; transgenic Brassica plant; transformation event;

KW male-sterility gene; PCR primer; thermal asymmetric interlaced;

KW TAIL; ss.

XX OS Bacillus amyloliquefaciens.

XX XX

PN WO200131042-A2.

XX XX

PD 03-MAY-2001.

XX XX

PF 26-OCT-2000; 2000WO-EPI0680.

XX XX

PR 29-OCT-1999; 99US-0430497.

XX XX

PA (AVET ) AVENTIS CROPS SCIENCE NV.

XX Weston B, De Beuckeleer M;

PI Oligonucleotide #1

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by

PT harboring specific transformation events, particularly by presence of

PT male-sterility gene, at specific location in its genome -

XX

PS Example 3; Page 28; 53pp; English.  
 CC The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is primary thermal interlaced (TAIL)-PCR primer  
 CC MD98 used to left (3') border flanking region of elite event MS-B2.  
 CC This primer corresponds to position 5249-5272 of plasmid pFColl3.  
 XX SQ Sequence 24 BP; 7 A; 9 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 24; DB 22; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 0.036;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 1 tcagaagtatcagcgacctccacc 24

RESULT 2  
 AAV23236  
 ID AAV23236 standard; DNA; 270 BP.  
 XX AC AAV23236;  
 XX DT 17-JUL-1998 (first entry)  
 XX DE Wild type barstar DNA.  
 XX KW Barstar; barnase inhibitor; fertility restoration;  
 XX KW male-sterile line; ds.  
 XX OS Bacillus amyloliquefaciens.  
 XX FH Key Location/Qualifiers  
 XX FT CDS 1..273  
 XX FT /\*tag= a  
 XX FT /product= barstar  
 XX FT /note= "stop codon not given"  
 XX PN WO9810081-A2.  
 XX PD 12-MAR-1998.  
 XX PF 01-SEP-1997; 97WO-EP04739.  
 XX PR 03-SEP-1996; 96EP-0202446.  
 XX PA (PLB2 ) PLANT GENETIC SYSTEMS NV.  
 XX PI Michiels F, Williams M;  
 XX WPI: 1998-193630/17.  
 XX P-PSDB; AAW53344.  
 XX DNA encoding an improved barstar protein - used to restore fertility  
 XX in male-sterile plant lines  
 XX Claim 8; Pages 34-35; 54pp; English.  
 XX The present sequence was used in the preparation of an improved  
 XX Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which  
 XX can be used to restore fertility to male-sterile lines.  
 XX The DNA sequence encoding the improved barstar, leads to increased  
 XX barstar production in tapetum cells, due to improved translation,  
 XX and possibly protein stability.

SQ Sequence 270 BP; 80 A; 58 C; 74 G; 58 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 0.048;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 32 tcagaagtatcagcgacctccacc 55

RESULT 3  
 AAV60977  
 ID AAV60977 standard; DNA; 323 BP.  
 XX AC AAV60977;  
 XX DT 03-DEC-1998 (first entry)  
 XX DE Barstar coding sequence.  
 XX KW Barnase; barstar; IPCR; inverse polymerase chain reaction; phenotype;  
 XX KW transgenic plant; hybrid seed; male sterile plant; active enzyme;  
 XX KW regulatory protein; embryoless seed; herbicide resistance; ss.  
 XX OS Synthetic.

FH Key Location/Qualifiers  
 FT CDS 1..273  
 FT /\*tag= a  
 FT /product= "barstar"

PN WO9837211-A1.

XX PD 27-AUG-1998.

XX PF 20-FEB-1998; 98WO-GB00542.

XX PR 21-FEB-1997; 97GB-0003681.

XX PA (GENE-) GENE SHEARS PTY LTD.

XX PI Betzner AS, Huttner E, Paul W, Perez P;

XX WPI: 1998-467572/40.

XX P-PSDB; AAW71704.

XX Production of transgenic plants having a desired phenotype - by  
 XX using a pair of parent plants which each produce a polypeptide which  
 XX complement each other when crossed

XX Example 1; Fig 1C; 58pp; English.

XX The present invention describes a pair of parent plants for producing  
 XX seeds comprising: (a) a first parent plant containing at least 1 gene  
 XX sequence encoding a polypeptide or protein A, and (b) a second parent  
 XX plant containing at least 1 gene sequence encoding a polypeptide or  
 XX protein B; where the polypeptides A and B, when expressed in separate  
 XX plants, do not form an active enzyme, a regulatory protein or protein  
 XX which affects the functionality and/or viability and/or the structural  
 XX integrity of a cell, but when expressed in the same plant do form an  
 XX active enzyme, regulatory protein, or protein which affects the  
 XX structural integrity of a cell. Also described is a method for producing  
 XX a plant having a desired phenotype by virtue of an active enzyme, a  
 XX regulatory protein or a protein which affects the structural integrity  
 XX of a cell comprising crossing a first line with a second line where the  
 XX first line contains one or more gene sequences encoding a polypeptide or  
 XX protein A but which line does not have the desired phenotype and where  
 XX the second line contains one or more gene sequences encoding a  
 XX polypeptide or protein B which is complementary to the polypeptide or  
 XX protein A but which line does not have the desired phenotype. The method  
 XX can be used for producing plants having altered phenotypes, e.g. male-

CC sterility, embryoless seeds, altered biochemical (e.g. fatty acid)  
 CC composition or herbicide resistance. The present sequence encodes  
 CC barstar which is used in an example from the present invention.

XX Sequence 323 BP; 98 A; 69 C; 87 G; 69 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 0.049;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcagaagtacagcagctccacc 24  
 |||||  
 Db 32 tcagaagtacagcagctccacc 55

RESULT 4  
 AAQ10460  
 ID AAQ10460 standard; DNA; 340 BP.  
 XX AC AAQ10460;  
 XX DT 16-APR-1991 (first entry)  
 XX DE ClaI-HindII fragment contg barstar gene.

XX Inhibitor; extracellular ribonuclease; barnase; fertility; ss.

XX Bacillus amyloliquefaciens.

XX EP412911-A.

XX 13-FEB-1991.

XX 09-AUG-1990; 90EP-0402281.

XX 10-AUG-1989; 89EP-0402270.

XX (PLAN-) PLANT GENETIC SYST.

XX Mariani C, Leemans J, De Greef W;

XX WPI; 1991-046026/07.

XX Cell of fertility restored plant - in which nuclear genome in  
 PT transformed with foreign DNA sequence neutralising activity of  
 PT another prod.

XX Disclosure; Fig 2; 25pp; English.

XX The barstar gene encodes and inhibitor of barnase, which degrades  
 CC RNA molecules by hybridising the bond after a guanine residue.  
 CC The gene is used, in a chimaeric sequence, to restore fertility in  
 CC plants transformed with the gene. The promoter cassette PTA29  
 CC (EP-401194) is fused in frame with the initiating AUG.

XX Sequence 340 BP; 106 A; 73 C; 92 G; 69 T; 0 other;

Query Match 100.0%; Score 24; DB 12; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 0.05;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcagaagtacagcagctccacc 24  
 |||||  
 Db 42 tcagaagtacagcagctccacc 65

RESULT 5  
 AAT39337  
 ID AAT39337 standard; DNA; 1303 BP.  
 XX AC AAT39337;

XX 22-JAN-1997 (first entry)  
 XX Plasmid pTS88 (EcoRI-HindIII fragment).  
 XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;  
 KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.  
 XX Synthetic.  
 XX Key Location/Qualifiers  
 FT misc\_feature 1..35  
 FT /\*tag= a  
 FT /label= pGEM2  
 FT /note= "polylinker of pGEM2"  
 FT 36..694  
 FT /\*tag= b  
 FT /label= P35S  
 FT /function= 35S promoter of cauliflower mosaic virus  
 FT strain CM1841  
 FT CDS 695..967  
 FT /\*tag= c  
 FT /label= barstar  
 FT /product= Bacillus amyloliquefaciens barstar  
 FT 968..1287  
 FT /\*tag= d  
 FT /label= 3'g7  
 FT /function= region containing polyadenylation signal  
 FT of gene 7 og Agrobacterium T-DNA  
 FT misc\_feature 1288..1303  
 FT /\*tag= e  
 FT /label= pGEM2  
 FT /note= "polylinker of pGEM2"  
 XX WO9626283-A1.  
 XX 29-AUG-1996.  
 XX 21-FEB-1996; 96WO-EP00722.  
 XX 21-FEB-1995; 95EP-0400364.  
 XX (PLBZ ) PLANT GENETIC SYSTEMS NV.  
 XX Botterman J, Cornelissen M, Michiels F;  
 XX WPI; 1996-402373/40.  
 XX Prodn. of male sterile plants by transforming with a chimaeric  
 PT construct - comprising a male sterility DNA e.g. barnase and a  
 PT co-regulating gene, e.g. barstar, into the nuclear genome, useful  
 PT for generating hybrid cultivars  
 XX Example 1; Page 38; 56pp; English.  
 XX The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains  
 CC barstar DNA under control of a 35S promoter. The plasmid was  
 CC used with pTS174 (see also AAT39336) contg. barnase DNA under  
 CC control of the stamen-specific promoter El to produce male sterile  
 CC rice cv. Kochihibiki transgenic plants, and with plasmid pVE136  
 CC (see also AAT39338) contg. barnase DNA under control of the stamen-  
 CC specific PCA55 promoter to produce male sterile maize plants.  
 CC Expression of barnase (a ribonuclease) in the stamen leads to male  
 CC sterility. Constitutive expression of barstar counteracts possible  
 CC low level expression of barnase DNA in non-stamen tissue.  
 XX Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 100.0%; Score 24; DB 17; Length 1303;  
 Best Local Similarity 100.0%; Pred. No. 0.059;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

QY 1 tcagaagtatcagcgacctccacc 24
   |||||||
Db 594 TCAGAAGTATCAGCGACCTCCACC 571

RESULT 8
AA15632/c
ID AAX15632 standard; DNA; 3544 BP.
XX
AC AAX15632;
XX
DT 07-MAY-1999 (first entry)
XX
DE Nucleotide sequence of an EcoRI-HindIII fragment of plasmid pTS256.
XX
KW Cl gene; maize; male-sterile corn line; anthocyanin production;
KW pTS256; ds.
XX
OS Synthetic.
XX
PN US5880331-A.
XX
PD 09-MAR-1999.
XX
PF 07-JUN-1995; 95US-0485139.
XX
PR 07-JUN-1995; 95US-0485139.
XX
PR 06-JUN-1994; 94US-0254776.
XX
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
PI Krebbers E, Leemans J, Williams M;
XX
DR WPI; 1999-204053/17.
XX
PT Process for maintaining a male-sterile corn line - using male
PT sterile plants lacking functional regulatory gene for anthocyanin
PT production
XX
PS Example 2; Columns 37-40; 35pp; English.
XX
CC The present sequence represents the nucleotide sequence of an
CC EcoRI-hindIII fragment of plasmid pTS256, comprising the chimeric
CC gene PCA55-barstar-3'nos. It is used in the course of the invention. The
CC specification describes a process for maintaining a male-sterile corn
CC line, using male sterile parent plants lacking a functional gene for
CC anthocyanin production, and a maintainer corn line comprising male
CC fertile parent plants containing foreign DNA comprising a restorer gene
CC and an active regulatory protein gene. By using the anthocyanin gene,
CC the colour of the male-sterile plants will differ from that of the
CC male fertile plants. This will enable the seeds harvested from the
CC plants to be easily separated into those that will grow into
CC male-fertile plants and those that will grow into male-sterile plants.
XX
SQ Sequence 3544 BP; 1011 A; 847 C; 767 G; 919 T; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 3544;
Best Local Similarity 100.0%; Pred. NO. 0.067;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24
   |||||||
Db 594 TCAGAAGTATCAGCGACCTCCACC 571

RESULT 9
AAV23237
ID AAV23237 standard; DNA; 4032 BP.
XX
AC AAV23237;
XX

```

```

DT 17-JUL-1998 (first entry)
XX
DE Plasmid pmw71.
XX
KW Barstar; barnase inhibitor; fertility restoration;
KW male-sterile line; plasmid pmw71; circular; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT 1999..3400
FT /*tag= a
FT /note= "promoter region of rice actin gene -
FT contains an intron in the leader"
FT CDS 3401..3676
FT /*tag= b
FT /product= barstar
FT 3'UTR 3677..4003
FT /*tag= c
FT /note= "region containing 3' untranslated end of
FT the nopaline synthase gene of Agrobacterium
FT T-DNA"
XX
PN WO9810081-A2.
XX
PD 12-MAR-1998.
XX
PF 01-SEP-1997; 97WO-EP04739.
XX
PR 03-SEP-1996; 96EP-0202446.
XX
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
PI Michiels F, Williams M;
XX
DR WPI; 1998-193630/17.
XX
PT DNA encoding an improved barstar protein - used to restore fertility
PT in male-sterile plant lines
XX
PS Example 3; Pages 37-39; 54pp; English.
XX
CC The present sequence was used in the preparation of an improved
CC Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which
CC can be used to restore fertility to male-sterile lines.
CC The DNA sequence encoding the improved barstar, leads to increased
CC barstar production in tapetum cells, due to improved translation,
CC and possibly protein stability.
XX
SQ Sequence 4032 BP; 1072 A; 968 C; 963 G; 1029 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 4032;
Best Local Similarity 100.0%; Pred. NO. 0.068;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24
   |||||||
Db 3435 tcagaagtatcagcgacctccacc 3458

RESULT 10
AAQ53889/c
ID AAQ53889 standard; DNA; 4808 BP.
XX
AC AAQ53889;
XX
DT 27-JUN-1994 (first entry)
XX
DE Restriction fragment of construct carrying plant maintainer gene.
XX Maintainer gene; sterile; sterility; homogenous population; hybrid;
KW seed; fertility restorer gene; pollen lethality gene; ss.

```

OS Synthetic.  
 XX Key  
 FH misc\_signal  
 FT Location/Qualifiers  
 FT complement (18..401)  
 FT /tag= a  
 FT /label= 3' nos.  
 FT /note= "3' regulatory sequence containing the  
 FT polyadenylation site derived from  
 FT Agrobacterium T-DNA nopaline synthase gene."  
 FT complement (402..737)  
 FT /tag= b  
 FT /label= Barnase.  
 FT /note= "Coding region of the barnase gene of  
 FT Bacillus amyloliquefaciens."  
 FT complement (738..1944)  
 FT /tag= c  
 FT /label= P2M13.  
 FT /note= "Promoter region of the Zm 13 gene of Zea  
 FT mays."  
 FT complement (1945..2281)  
 FT /tag= d  
 FT /label= 3' nos.  
 FT complement (2282..2554)  
 FT /tag= e  
 FT /label= Barstar.  
 FT /note= "Coding region of the barstar gene of  
 FT Bacillus amyloliquefaciens."  
 FT complement (2555..3099)  
 FT /tag= f  
 FT /label= PTA29.  
 FT /note= "Promoter region of the TA29 gene of  
 FT Nicotiana tabacum."  
 FT 3100..3932  
 FT /tag= g  
 FT /label= 35S3.  
 FT /note= "35S3" promoter sequence derived from  
 FT cauliflower mosaic virus isolate CabbB-JI."  
 FT 3933..4484  
 FT /tag= h  
 FT /note= "Coding region of the phoshinothricin  
 FT acetyltransferase gene."  
 FT 4485..4763  
 FT /tag= i  
 FT /label= 3' nos.  
 FT  
 XX WO9325695-A.  
 XX  
 XX 23-DEC-1993.  
 XX  
 XX 11-JUN-1993; 93WO-EP01489.  
 XX  
 XX 12-JUN-1992; 92US-0899072.  
 PR 03-NOV-1992; 92US-0970840.  
 XX  
 XX (PLB2 ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Leemans J, Williams M;  
 XX WPI; 1994-007552/01.  
 XX  
 XX Maintainer gene for maintenance of male-sterile plants -  
 PT comprises fertility-restorer gene and pollen-lethality gene  
 XX  
 PS Example 2; Page 73-75; 87pp; English.  
 XX  
 CC A maintainer gene of plants, pref. a foreign chimeric gene,  
 CC comprises (a) a fertility restorer gene which comprises a fertility  
 CC restorer DNA and (ii) a restorer promoter capable of  
 CC directing the expression of the fertility restorer DNA and (b) a  
 CC pollen lethality gene that is selectively expressed in microspores  
 CC and/or pollen of the plant to prevent the production of functional  
 CC pollen and which comprises (i) a pollen lethality DNA and (ii) a

CC pollen specific promoter capable of directing expression of the  
 CC pollen lethality DNA. Plants transformed with this DNA (maintainer  
 CC plants) can be used to maintain a homogeneous population of male  
 CC sterile plants for the production of hybrid seed. This sequence is  
 CC an EcoRI-HindIII restriction fragment of the construct designated  
 CC pTS218 and comprises the maintainer gene described.  
 XX  
 SQ Sequence 4808 BP; 1370 A; 1064 C; 1037 G; 1333 T; 4 other;  
 Query Match 100.0%; Score 24; DB 15; Length 4808;  
 Best Local Similarity 100.0%; Pred. No. 0.069; 0; Gaps 0;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0;  
 QY 1 tcagaagatcagcgacctccacc 24  
 Db 2529 TCAGAGTATCAGCGACCTCCACC 2506  
 RESULT 11  
 AAH25423/C  
 ID AAH25423 standard; DNA; 4832 BP.  
 XX  
 AC AAH25423;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 DE Nucleotide sequence of plasmid pTHW118.  
 XX  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barstar gene; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 OS Arabidopsis thaliana.  
 OS Bacillus amyloliquefaciens.  
 OS Nicotiana tabacum.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..25  
 FT /tag= a  
 FT /note= "right border repeat from TL-DNA from pT1B6S3"  
 FT 26..53  
 FT /tag= b  
 FT /note= "synthetic polylinker derived sequences"  
 FT 54..90  
 FT /tag= c  
 FT /note= "residual sequence from TL-DNA at right  
 FT border repeat"  
 FT 91..97  
 FT /tag= d  
 FT /note= "synthetic polylinker derived sequences"  
 FT complement (98..309)  
 FT /tag= e  
 FT /note= "3' UTR from TL-DNA gene 7 of pT1B6S3"  
 FT 310..330  
 FT /tag= f  
 FT /note= "synthetic polylinker derived sequences"  
 FT complement (331..882)  
 FT /tag= g  
 FT /note= "Streptomyces hygroscopicus bialaphos  
 FT resistance (bar) gene"  
 FT complement (883..2608)  
 FT /tag= h  
 FT /note= "atsIA ribulose-1,5-biphosphate carboxylase  
 FT small subunit gene from Arabidopsis thaliana"  
 FT 2609..2658  
 FT /tag= i  
 FT /note= "synthetic polylinker derived sequences"  
 FT complement (2659..2919)  
 FT /tag= j  
 FT /note= "TaqI fragment from 3' UTR of nopaline  
 FT synthase gene from T-DNA of pTIT37 and



```

FT misc_feature 2920..2940 containing plant polyadenylation signals"
FT /*tag= k
FT /note= "synthetic polylinker derived sequences"
FT 2941..2980
FT /*tag= l
FT /note= "downstream of Bacillus amyloliquefaciens
FT barstar coding region"
FT CDS complement (2981..3253)
FT /*tag= m
FT /note= "Barstar gene coding region from Bacillus
FT amyloliquefaciens"
FT promoter complement (3254..4762)
FT /*tag= n
FT /note= "anther-specific gene TA29 promoter from
FT Nicotiana tabacum"
FT misc_feature 4763..4807
FT /*tag= o
FT /note= "synthetic polylinker derived sequences"
FT 4808..4832
FT /*tag= p
FT /note= "left border repeat from TL-DNA from pTiB6S3"
FT WO200141558-A1.
PN
XX
XX 14-JUN-2001.
PD
XX
XX 06-DEC-2000; 2000WO-EPI2872.
PF
XX
XX 08-DEC-1999; 99US-0457037.
PR
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
PA
XX
XX De Both G, De Beuckeleer M;
PI
XX
XX WPT; 2001-381419/40.
DR
XX
XX Transgenic winter oilseed rape plants suited for producing hybrid seed
FT with improved qualities, comprises a male-sterility gene and fertility
FT restorer gene, integrated into the genome
FT
XX Example 1; Page 80-82; 98pp; English.
XX
XX The specification describes a pair of transgenic winter oilseed rape
CC plants suited for producing hybrid seed. One of the plants has an
CC expression cassette comprising a male-sterility gene and the other
CC plant has an expression cassette comprising a fertility restorer gene,
CC integrated into the genome. The fertility restorer gene is capable of
CC preventing the activity of the male-sterility gene. The plant pair is
CC useful for producing hybrid seed. Plants developed from the hybrid
CC seed have agronomic performance, genetic stability and adaptability to
CC different genetic backgrounds. The present sequence represents
CC plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
CC a fertility restorer gene. The plasmid is used to create transgenic
CC plants of the invention.
XX
XX Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;
SQ

```

```

Query Match 100.0%; Score 24; DB 22; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 tcagaagtatcagcgacctccacc 24
|||||
Db 3222 TCAGAAGTATCAGCGACCTCCACC 3199

```

RESULT 12

AA08976

ID AA08976 standard; DNA; 4896 BP.

XX

AC AA08976;

```

XX 12-AUG-1996 (first entry)
XX
XX pTS256 contg. P35S-bar-3'nos and PTA29-barstar-3'nos.
XX
XX C1; C1-S; transgenic plant; male sterility; colour-linked restorer;
XX anthocyanin; aleurone; B-peru; bar; barstar; bar*; ds.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX misc_signal complement (39..317)
XX /*tag= a
XX /label= 3'nos
XX /note= "3' regulatory sequence contg. the
XX polyadenylation signal of the nopaline
XX synthase gene of Agrobacterium T-DNA"
XX CDS complement (318..869)
XX /*tag= b
XX /label= bar
XX /note= "coding region of bar gene of
XX Streptomyces hygroscopicus"
XX promoter complement (870..1702)
XX /*tag= c
XX /label= P35S
XX /note= "35S promoter of Cauliflower Mosaic Virus"
XX complement (1740..2284)
XX /*tag= d
XX /label= PTA29
XX /note= "promoter of TA29 gene of Nicotiana
XX tabacum"
XX CDS 2285..2557
XX /*tag= e
XX /label= barstar
XX /note= "coding region of barstar gene of
XX Bacillus amyloliquefaciens"
XX misc_signal 2558..2879
XX /*tag= f
XX /label= 3'nos
XX /note= "3' regulatory sequence contg. the
XX polyadenylation signal of the nopaline
XX synthase gene of Agrobacterium T-DNA"
XX
XX misc_RNA 1..38
XX /*tag= g
XX /label= pUC19
XX /note= "pUC19 derived sequence"
XX 2880..4986
XX /*tag= h
XX /label= pUC19
XX /note= "pUC19 derived sequence"
XX
XX WO9534634-A2.
XX
XX 21-DEC-1995.
XX
XX 06-JUN-1995; 95WO-EP02157.
XX
XX 06-JUN-1994; 94US-0254776.
XX (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
XX Krebbers E, Leemans J, Williams M;
XX
XX WPI; 1996-049664/05.
XX
XX Transgenic plants contg. male sterility and colour-linked restorer
XX genotypes - used for prodn. of male sterile seeds identifiable from
XX their colour, also new truncated anthocyanin regulatory genes and
XX aleurone specific promoters
XX
XX Disclosure; Page 68-71; 104pp; English.
XX
XX Shortened forms of anthocyanin (An) regulatory genes (ARG) are easier
XX

```

CC to manipulate than the complete gene and still provide An prodn.  
 CC The C1 gene (and the C1-S gene) can be considerably shortened  
 CC while still retaining, under appropriate conditions, its  
 CC capability of conditioning anthocyanin prodn. in the aleurone  
 CC of seeds of cereal plants such as corn. A pref. shortened C1 gene  
 CC is comprised in pCOL9 (see AAT08975). The full C1 gene sequence is  
 CC given in AAT08973.  
 CC A truncated B-peru gene (presumed sequence = AAT08674; actual  
 CC sequence = AAT08977) may also be used.  
 CC Plasmids pTS256 (AAT08976) and pTS200 (AAT17246) were used in the  
 CC construction of vectors comprising the C1 and B-peru genes  
 CC as well as male-sterility gene and a selectable marker gene.  
 XX  
 SQ Sequence 4896 BP; 1252 A; 1146 C; 1186 G; 1312 T; 0 other;

Query Match 100.0%; Score 24; DB 17; Length 4896;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 2316 tcagaagtatcagcgacctccacc 2339

RESULT 13

ID AAX15631 standard; DNA; 4896 BP.

AC AAX15631;

DT 07-MAY-1999 (first entry)

DE Nucleotide sequence of plasmid pTS256.

KW C1 gene; maize; male-sterile corn line; anthocyanin production;  
 KW pTS256; ds.

OS Synthetic.

PN US5880331-A.

PD 09-MAR-1999.

PF 07-JUN-1995; 95US-0485139.

PR 07-JUN-1995; 95US-0485139.

PR 06-JUN-1994; 94US-0254776.

PA (PLBZ ) PLANT GENETIC SYSTEMS NV.

PI Krebbers E, Leemans J, Williams M;

DR WPI; 1999-204053/17.

XX Process for maintaining a male-sterile corn line - using male  
 PT sterile plants lacking functional regulatory gene for anthocyanin  
 PT production

XX Example 2; Columns 31-36; 35pp; English.

XX The present sequence represents the nucleotide sequence of plasmid  
 CC pTS256. It is used in the course of the invention. The specification  
 CC describes a process for maintaining a male-sterile corn line, using male  
 CC sterile parent plants lacking a functional gene for anthocyanin  
 CC production, and a maintainer corn line comprising male fertile parent  
 CC plants containing foreign DNA comprising a restorer gene and an active  
 CC regulatory protein gene. By using the anthocyanin gene, the colour of  
 CC the male-sterile plants will differ from that of the male fertile  
 CC plants. This will enable the seeds harvested from the plants to be  
 CC easily separated into those that will grow into male-fertile plants and  
 CC those that will grow into male-sterile plants.

XX

SQ Sequence 4896 BP; 1251 A; 1147 C; 1186 G; 1312 T; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 4896;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 2317 tcagaagtatcagcgacctccacc 2340

RESULT 14

AAT39339

ID AAT39339 standard; DNA; 5864 BP.

XX AAT39339;

AC AAT39339;

DT 22-JAN-1997 (first entry)

XX Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.  
 XX Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;  
 KW transgenic plant; oilseed rape; canole; Brassica napus; ds.  
 XX Synthetic.

OS

EH Key Location/Qualifiers  
 FT misc\_feature complement (1..25)  
 FT /tag= a

FT /label= RB

FT /note= "right border of Agrobacterium T-DNA"

FT complement (98..330)

FT /tag= b

FT /label= 3'g7

FT /note= "region containing polyA signal of gene 7

FT Of Agrobacterium T-DNA"

FT complement (331..882)

FT /tag= c

FT /label= bar

FT /note= "region coding for phosphinothricin

FT acetyltransferase"

FT complement (883..2608)

FT /tag= d

FT /label= Pssu

FT /note= "promoter of Arabidopsis Rubisco small

FT subunit gene"

FT complement (2659..3031)

FT /tag= e

FT /label= 3'nos

FT /note= "region containing polyA signal of nopaline

FT synthase gene of Agrobacterium T-DNA"

FT complement (3032..3367)

FT /tag= f

FT /label= Barnase

FT /note= "Bacillus amyloliquefaciens barnase coding

FT region"

FT complement (3368..4877)

FT /tag= g

FT /label= pTA29

FT /note= "promoter of stamen-specific TA29 gene of

FT 4924..5216

FT /tag= h

FT /label= Pnos

FT /note= "promoter of nopaline synthase gene of

FT Agrobacterium T-DNA"

FT 5217..5489

FT /tag= i

FT /label= Barstar

FT /note= "region coding for barstar of Bacillus

FT amyloliquefaciens"

FT 5490..5765

FT polyA\_signal

FT	/*tag= j		
FT	/label= 3'g7		
FT	/note= "region containing polyA signal of gene 7		
FT	of Agrobacterium T-DNA"		
FT	misc_feature	complement (5840..5864)	
FT	/*tag= k		
FT	/label= LB		
FT	/note= "left border of Agrobacterium T-DNA"		
XX			
PN	W09626283-A1.		
XX			
PD	29-AUG-1996.		
XX			
PF	21-FEB-1996;	96WO-EP00722.	
XX			
PR	21-FEB-1995;	95EP-0400364.	
XX			
PA	(PLB2 ) PLANT GENETIC SYSTEMS NV.		
XX			
PI	Botterman J, Cornelissen M, Michiels F;		
XX	WPI; 1996-402373/40.		
DR			
XX			
XX	Prodn. of male sterile plants by transforming with a chimaeric		
PT	construct - comprising a male sterility DNA e.g. barnase and a		
PT	co-regulating gene, e.g. barstar, into the nuclear genome, useful		
PT	for generating hybrid cultivars		
XX			
PS	Example 3; Page 33-3743-47; 56pp; English.		
XX			
CC	Plasmid pTColl3 (AAT39339) is a T-DNA vector containing a bar gene		
CC	under control of the PSSU promoter, a barnase gene under control		
CC	of the stamen-specific PRA29 promoter, and a barstar (co-regulatory)		
CC	gene under control of the pnos promoter. 87% Of oilseed rape		
CC	plants regenerated after Agrobacterium-mediated transformation		
CC	using pTColl3 were male sterile. Barnase expression disturbed the		
CC	function of stamen cells leading to male sterility. Constitutive		
CC	expression of barstar counteracted any low level expression of		
CC	barnase in non-stamen tissue.		
XX			
SQ	Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;		
	Query Match 100.0%; Score 24; DB 17; Length 5864;		
	Best Local Similarity 100.0%; Pred. NO. 0.071;		
	Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 tcagaagtatcagcgactccacc 24		
Db	5248 tcagaagtatcagcgactccacc 5271		
	RESULT 15		
AAD06990			
ID	AAD06990 standard; DNA; 5865 BP.		
XX			
AC	AAD06990;		
XX			
DT	06-AUG-2001 (first entry)		
XX			
DE	Chimeric T-DNA of plasmid pTColl3.		
XX			
XX	T-DNA; plasmid pTColl3; transgenic Brassica plant; transformation event;		
KW	male-sterility gene; chimeric; tobacco; ds.		
XX			
OS	Chimeric - Streptomyces hygroscopicus.		
OS	Chimeric - Arabidopsis thaliana.		
OS	Chimeric - Bacillus amyloliquefaciens.		
OS	Chimeric - Nicotiana tabacum.		
OS	Chimeric - Agrobacterium tumefaciens.		
OS	Chimeric - Unidentified.		
XX			
FH	Key Location/Qualifiers		

FT	misc_feature	1..25	
FT	/*tag= a		
FT	/note= "Right border repeat from the TL-DNA from		
FT	pTiB6S3"		
FT	26..53		
FT	/*tag= b		
FT	/note= "Synthetic polylinker derived sequence"		
FT	54..90		
FT	/*tag= c		
FT	/note= "Residual sequence from the TL-DNA at the		
FT	right border repeat"		
FT	98..309		
FT	/*tag= d		
FT	/note= "The 3' untranslated end from the TL-DNA		
FT	gene 7 (3'g7) of pTiB6S3"		
FT	310..331		
FT	/*tag= e		
FT	/note= "Synthetic polylinker derived sequence"		
FT	332..883		
FT	/*tag= f		
FT	/product= "Protein encoded by bialaphos resistance		
FT	gene (bar) of Streptomyces hygroscopicus"		
FT	884..2609		
FT	/*tag= g		
FT	/note= "Promoter from the atsIA ribulose-1,5-biphosphate		
FT	carboxylase small subunit gene from Arabidopsis thaliana"		
FT	2610..2659		
FT	/*tag= h		
FT	/note= "Synthetic polylinker derived sequence"		
FT	2660..2920		
FT	/*tag= i		
FT	/note= "TaqI fragment from the 3' untranslated end of the		
FT	nopaline synthase gene (3'nos) from the T-DNA of pTiT37		
FT	and containing plant polyadenylation signals"		
FT	2921..2936		
FT	/*tag= j		
FT	/note= "Synthetic polylinker derived sequence"		
FT	2937..3032		
FT	/*tag= k		
FT	/note= "The 3' untranslated region downstream from the		
FT	barnase coding sequence of Bacillus amyloliquefaciens"		
FT	3033..3368		
FT	/*tag= l		
FT	/product= "Protein encoded by barnase gene from		
FT	Bacillus amyloliquefaciens"		
FT	3369..4878		
FT	/*tag= m		
FT	/note= "Promoter region of the anther-specific gene		
FT	TA29 from Nicotiana tabacum"		
FT	4879..4924		
FT	/*tag= n		
FT	/note= "Synthetic polylinker derived sequence"		
FT	4925..5215		
FT	/*tag= o		
FT	/note= "Promoter of the nopaline synthase gene from the		
FT	T-DNA of pTiT37 of Agrobacterium tumefaciens"		
FT	5216..5217		
FT	/*tag= p		
FT	/note= "Synthetic polylinker derived sequence"		
FT	5218..5490		
FT	/*tag= q		
FT	/product= "Protein encoded by barstar gene of		
FT	Bacillus amyloliquefaciens"		
FT	5491..5530		
FT	/*tag= r		
FT	/note= "Sequence from the 3' untranslated end of the		
FT	barstar gene from Bacillus amyloliquefaciens"		
FT	5531..5554		
FT	/*tag= s		
FT	/note= "Synthetic polylinker derived sequence"		
FT	5555..5766		
FT	/*tag= t		
FT	/note= "The 3' untranslated end from the TL-DNA		

```
FT misc_feature      gene 7 (3'g7) of pTiB6S3"
FT 5767..5773
FT /*tag= u
FT /note= "Synthetic polylinker derived sequence"
FT 5774..5810
FT /*tag= v
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 5811..5840
FT /*tag= w
FT /note= "Synthetic polylinker derived sequence"
FT 5841..5865
FT /*tag= x
FT /note= "Left border repeat from the TL-DNA from
FT pTiB6S3"
XX
PN WO200131042-A2.
XX
XX 03-MAY-2001.
XX
XX 26-OCT-2000; 200WO-EP10680.
XX
XX 29-OCT-1999; 99US-0430497.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX Weston B, De Beuckeleer M;
XX
XX WPI; 2001-300517/31.
XX
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX harboring specific transformation events, particularly by presence of
XX male-sterility gene, at specific location in its genome -
XX
XX Claim 1; Page 47-49; 53pp; English.
XX
XX The present invention relates to a transgenic Brassica plant or its
XX seed, cells or tissues, characterised by harbouring a specific
XX transformation event, particularly by the presence of a male-sterility
XX gene, at a specific location in the Brassica genome. Transgenic
XX Brassica plant is useful for producing a hybrid seed by crossing the
XX transgenic plant with a male-fertile Brassica plant and harvesting the
XX hybrid seed from the transgenic Brassica plant.
XX The present sequence is chimeric T-DNA of plasmid pTCO113. This sequence
XX comprises right border repeat, left border repeat and 3' untranslated
XX region (UTR) from TL-DNA of pTiB6S3, synthetic polylinker sequences,
XX coding regions of biolaphos resistance gene (bar) from
XX Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens
XX and barstar gene from Bacillus amyloliquefaciens and promoters of atS1A
XX ribulose-1,5-bisphosphate carboxylase small subunit gene from
XX Arabidopsis thaliana, the anther-specific gene TA29 from
XX Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTiT37
XX of Agrobacterium tumefaciens.
XX
XX Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;
```

Query Match 100.0%; Score 24; DB 22; Length 5865;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 tcagaagtatcagcgacctccacc 24
   |||
Db 5249 tcagaagtatcagcgacctccacc 5272
```

Search completed: February 15, 2002, 19:01:55  
Job time: 20759 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 19:08:07 ; Search time 353.79 Seconds  
(without alignments)  
15.364 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24  
Sequence: 1 tcgaagtatcagcgacctccacc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	1303	3	US-08-894-440-2
C 2	24	100.0	3544	2	US-08-485-139-3
C 3	24	100.0	3544	3	US-08-750-357-3
C 4	24	100.0	4808	1	US-08-351-413-17
C 5	24	100.0	4808	2	US-09-025-583-17
C 6	24	100.0	4896	2	US-08-485-139-2
C 7	24	100.0	4896	3	US-08-750-357-2
C 8	24	100.0	5864	3	US-08-894-440-4
C 9	24	100.0	6555	1	US-08-351-413-2
C 10	24	100.0	6555	2	US-09-025-583-2
C 11	16.6	69.2	1866	1	US-08-403-388-1
C 12	16.6	69.2	1866	1	US-08-658-578-1
C 13	16.6	69.2	1866	3	US-08-846-111D-1
C 14	16.6	69.2	1866	4	US-09-056-105-19
C 15	16.6	69.2	40352	3	US-08-846-111D-15
C 16	16	66.7	1128	2	US-09-016-366A-20
C 17	16	66.7	1128	2	US-08-978-404B-15
C 18	16	66.7	1137	2	US-09-016-366A-18
C 19	16	66.7	1137	2	US-08-978-404B-13
C 20	16	66.7	1154	2	US-09-016-366A-16
C 21	16	66.7	1154	4	US-08-978-404B-11
C 22	15.6	65.0	4286	4	US-09-413-304-7
C 23	15.6	65.0	4437	1	US-08-559-303B-72
C 24	15.6	65.0	4437	4	US-09-175-828-72
C 25	15.6	65.0	6519	1	US-08-588-985-1
C 26	15.6	65.0	6519	1	US-08-971-988-1
C 27	15.2	63.3	363	1	US-08-594-031-104

c	28	15.2	63.3	1082	1	US-08-716-301-5	Sequence 5, Appli
	29	15.2	63.3	2326	2	US-08-231-193A-41	Sequence 41, Appl
	30	15.2	63.3	2326	2	US-08-486-273A-41	Sequence 41, Appl
	31	15.2	63.3	2326	3	US-08-480-474-41	Sequence 41, Appl
	32	15.2	63.3	2326	3	US-08-940-086A-41	Sequence 41, Appl
	33	15.2	63.3	3243	2	US-08-231-193A-44	Sequence 44, Appl
	34	15.2	63.3	3243	2	US-08-486-273A-44	Sequence 44, Appl
	35	15.2	63.3	3243	3	US-08-480-474-44	Sequence 44, Appl
	36	15.2	63.3	3243	3	US-08-940-086A-44	Sequence 44, Appl
	37	15.2	63.3	3698	2	US-08-231-193A-43	Sequence 43, Appl
	38	15.2	63.3	3698	2	US-08-486-273A-43	Sequence 43, Appl
	39	15.2	63.3	3698	3	US-08-480-474-43	Sequence 43, Appl
	40	15.2	63.3	3698	3	US-08-940-086A-43	Sequence 43, Appl
	41	15.2	63.3	3717	1	US-08-026-138E-7	Sequence 7, Appl
	42	15.2	63.3	4002	2	US-08-231-193A-53	Sequence 53, Appl
	43	15.2	63.3	4002	2	US-08-486-273A-53	Sequence 53, Appl
	44	15.2	63.3	4002	3	US-08-480-474-53	Sequence 53, Appl
	45	15.2	63.3	4002	3	US-08-940-086A-53	Sequence 53, Appl

ALIGNMENTS

RESULT 1  
US-08-894-440-2  
; Sequence 2, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894.440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1303  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI  
; OTHER INFORMATION: fragment of pTS88  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(35)  
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36)..(694)  
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain  
; OTHER INFORMATION: CM1841 (P35S)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (695)..(967)  
; OTHER INFORMATION: region coding for barstar of Bacillus  
; OTHER INFORMATION: amyloliquefaciens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (968)..(1287)  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1288)..(1303)  
; OTHER INFORMATION: polylinker of pGEM2  
; OTHER INFORMATION: polylinker of pGEM2  
; US-08-894-440-2

Query Match 100.0%; Score 24; DB 3; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 0.0042;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 tcgaagtatcagcgacctccacc 24

```

Db 726 tcagaagtatcagcagcccccacc 749
|||||
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of barstar gene of Bacillus
; OTHER INFORMATION: amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (30..352)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..6
; OTHER INFORMATION: /label= EcoRI
; FEATURE:
; NAME/KEY: -
; LOCATION: 3539..3544
; OTHER INFORMATION: /label= HindIII
; US-08-485-139-3

Query Match 100.0%; Score 24; DB 2; Length 3544;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcagcccccacc 24
|||||
Db 594 TCAGAAAGTATCAGCGACCTCCACC 571
|||||

RESULT 3
US-08-750-357-3/c
; Sequence 3, Application US/08750357
; Patent No. 6008437
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; TITLE OF INVENTION: MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750.357
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 018030-100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3544 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: EcoRI-HindIII region of plasmid pTS200
; FEATURE:
; NAME/KEY: -
; LOCATION: 3227..3504
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2675..3226
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces
; OTHER INFORMATION: hygroscopicus"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1841..2674
; OTHER INFORMATION: /label= p35S
; OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (626..1803)
; OTHER INFORMATION: /label= PCA55
; OTHER INFORMATION: /note= "promoter of CA55 gene of Zea mays"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (353..625)

```



```

; LOCATION: 3933..4484
; OTHER INFORMATION: /label= bar
; Best Local Similarity 100.0%; Score 24; DB 1; Length 4808;
; OTHER INFORMATION: /note= "coding region of the phosphinothricin
; FEATURE: acetyltransferase gene"
; NAME/KEY: -
; LOCATION: 4485..4763
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: 2333..2356
; OTHER INFORMATION: /label= BXOL2
; OTHER INFORMATION: /note= "region corresponding to oligonucleotide
; OTHER INFORMATION: BXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2538..2586)
; OTHER INFORMATION: /label= TA29SBXOL2
; OTHER INFORMATION: /note= "region complementary to oligonucleotide
; OTHER INFORMATION: TA29SBXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2800..2823)
; OTHER INFORMATION: /label= PTA29OL5
; OTHER INFORMATION: /note= "region complementary to part of
; OTHER INFORMATION: oligonucleotide PTA29OL5"
; US-08-351-413-17
;
; Query Match 100.0%; Score 24; DB 1; Length 4808;
; Best Local Similarity 100.0%; Pred. No. 0.0034;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
Qy 1 tcagaagtatcagcgacctccacc 24
    |||||||
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 5
; Sequence 17, Application US/09025583
; Patent No. 5977433
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,583
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; APPLICATION NUMBER: US 07/999,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.

```

```

; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: EcoRI-HindIII fragment of plasmid pTS218
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (18..401)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (402..737)
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding region of the barnase gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (738..1944)
; OTHER INFORMATION: /label= PZM13
; OTHER INFORMATION: /note= "promoter region of the Zml3 gene of Zea
; OTHER INFORMATION: mays"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (1945..2281)
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2282..2554)
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of the barstar gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2555..3099)
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "promoter region of the TA29 gene of
; OTHER INFORMATION: Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3100..3932
; OTHER INFORMATION: /label= 35S3
; OTHER INFORMATION: /note= "35S3" promoter sequence derived from
; OTHER INFORMATION: cauliflower mosaic virus isolate CabDB8-J1"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3933..4484
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of the phosphinothricin
; OTHER INFORMATION: acetyltransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4485..4763
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: 2333..2356
; OTHER INFORMATION: /label= BXOL2
; OTHER INFORMATION: /note= "region corresponding to oligonucleotide

```



```

; OTHER INFORMATION: BXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2538..2586)
; OTHER INFORMATION: /label= TA29SBXOL2
; OTHER INFORMATION: /note= "region complementary to oligonucleotide
; OTHER INFORMATION: TA29SBXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2800..2823)
; OTHER INFORMATION: /label= PTA29OL5
; OTHER INFORMATION: /note= "region complementary to part of
; OTHER INFORMATION: oligonucleotide PTA29OL5"
US-09-025-583-17
```

```

Query Match 100.0%; Score 24; DB 2; Length 4808;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 1 tcgaagtatcagcgacctccacc 24
|||||
Db 2529 TCAGAGTATCAGGACCTCCACC 2506
```

## RESULT 6

```

US-08-485-139-2
; Sequence 2, Application US/08485139
; Patent No. 5880331
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; TITLE OF INVENTION: MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,139
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid pTS256, linearized at HindIII
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (39..317)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
```

```

; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (318..869)
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces
; OTHER INFORMATION: hygrosopicus"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (870..1702)
; OTHER INFORMATION: /label= P35S
; OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1740..2284
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "promoter of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2285..2557
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of barstar gene of Bacillus
; OTHER INFORMATION: amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2558..2879
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..38
; OTHER INFORMATION: /label= pUC19
; OTHER INFORMATION: /note= "pUC19 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2880..4896
; OTHER INFORMATION: /label= pUC19
; OTHER INFORMATION: /note= "pUC19 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3004..3009
; OTHER INFORMATION: /label= EcoRI
US-08-485-139-2
```

```

Query Match 100.0%; Score 24; DB 2; Length 4896;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 1 tcgaagtatcagcgacctccacc 24
|||||
Db 2317 TCAGAGTATCAGGACCTCCACC 2340
```

## RESULT 7

```

US-08-750-357-2
; Sequence 2, Application US/08750357
; Patent No. 6008437
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; TITLE OF INVENTION: MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
```

STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/750,357  
FILING DATE: 21-MAR-1997  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 018030-100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4896 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: plasmid pTS256, linearized at HindIII  
FEATURE:  
NAME/KEY: -  
LOCATION: 39..317  
OTHER INFORMATION: /label= 3'nos  
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation  
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"  
FEATURE:  
NAME/KEY: -  
LOCATION: 318..869  
OTHER INFORMATION: /label= bar  
OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces hygroscopicus"  
FEATURE:  
NAME/KEY: -  
LOCATION: 870..1702  
OTHER INFORMATION: /label= p35S  
OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"  
FEATURE:  
NAME/KEY: -  
LOCATION: 1740..2284  
OTHER INFORMATION: /label= PTA29  
OTHER INFORMATION: /note= "promoter of TA29 gene of Nicotiana tabacum"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2285..2557  
OTHER INFORMATION: /label= barstar  
OTHER INFORMATION: /note= "coding region of barstar gene of  
OTHER INFORMATION: Bacillus amyloliquefaciens"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2558..2879  
OTHER INFORMATION: /label= 3'nos  
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation  
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..38  
OTHER INFORMATION: /label= pUC19  
OTHER INFORMATION: /note= "pUC19 derived sequence"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2880..4896  
OTHER INFORMATION: /label= pUC19  
OTHER INFORMATION: /note= "pUC19 derived sequence"  
FEATURE:  
NAME/KEY: -

LOCATION: 3004..3009  
OTHER INFORMATION: /label= EcoRI  
US-08-750-357-2

Query Match 100.0%; Score 24; DB 3; Length 4896;  
Best local similarity 100.0%; Pred. No. 0.0054;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgaagatcagcgacctccacc 24  
|||||  
Db 2316 TCAGAGTATCAGCGACCTCCACC 2339

## RESULT 8

US-08-894-440-4  
Sequence 4, Application US/08894440  
Patent No. 6025546  
GENERAL INFORMATION:  
APPLICANT: PLANT GENETIC SYSTEMS N.V.  
TITLE OF INVENTION: Method to obtain male sterile plants  
FILE REFERENCE: NMSCOR  
CURRENT APPLICATION NUMBER: US/08/894,440  
CURRENT FILING DATE: 1997-11-12  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 5864  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
OTHER INFORMATION: plasmid pTCO113  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((1)..(25))  
OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((98)..(330))  
OTHER INFORMATION: region containing polyadenylation signal of gene 7  
OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((331)..(882))  
OTHER INFORMATION: region coding for phosphinothricin acetyl  
OTHER INFORMATION: transferase (bar)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((883)..(2608))  
OTHER INFORMATION: promoter of small subunit gene of Rubisco of  
OTHER INFORMATION: Arabidopsis (Pssu)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((2659)..(3031))  
OTHER INFORMATION: region containing polyadenylation signal of  
OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((3032)..(3367))  
OTHER INFORMATION: region coding for barnase of Bacillus  
OTHER INFORMATION: amyloliquefaciens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((3368)..(4877))  
OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana  
OTHER INFORMATION: tabacum (PTA29)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (4924)..(5216)  
OTHER INFORMATION: promoter of nopaline synthase gene of  
OTHER INFORMATION: Agrobacterium T-DNA (Pnos)

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g')
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
; US-08-894-440-4

Query Match      100.0%; Score 24; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0056;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24
|||||

Db 5248 tcagaagtatcagcgacctccacc 5271

RESULT 9
US-08-351-413-2/c
; Sequence 2, Application US/08351413
; Patent No. 5750867
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6555 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
```

```
;
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: plasmid pVE144 (replicable in E.coli)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..396
; OTHER INFORMATION: /label= pUC18
; OTHER INFORMATION: /note= "pUC18 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (397..751)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (752..1024)
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of the barstar gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (1025..1607)
; OTHER INFORMATION: /label= TA29
; OTHER INFORMATION: /note= "promoter derived from the TA29 gene of
; OTHER INFORMATION: Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1608..2440
; OTHER INFORMATION: /label= 35S3
; OTHER INFORMATION: /note= "35S3 promoter sequence derived from
; OTHER INFORMATION: cauliflower mosaic virus isolate CabbbB-J1"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2441..3256
; OTHER INFORMATION: /label= neo
; OTHER INFORMATION: /note= "coding region of the neomycine
; OTHER INFORMATION: phosphotransferase gene of Tn5"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3257..4315
; OTHER INFORMATION: /label= 3'ocs
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA octopine synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4316..6555
; OTHER INFORMATION: /label= pUC18
; OTHER INFORMATION: /note= "pUC18 derived sequence"
; US-08-351-413-2

Query Match      100.0%; Score 24; DB 1; Length 6555;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24
|||||

Db 993 TCAGAAGTATCAGCGACCTCCACC 970

RESULT 10
US-09-025-583-2/c
; Sequence 2, Application US/09025583
; Patent No. 5977433
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
```

```
;; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
;; STREET: 8110 Gatehouse Road, Suite 500 East
;; CITY: Falls Church
;; STATE: Virginia
;; COUNTRY: U.S.A.
;; ZIP: 2046
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US 07/925,583
;; FILING DATE:
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/970,849
;; FILING DATE: 03-NOV-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Svensson, Leonard R.
;; REGISTRATION NUMBER: 30,330
;; REFERENCE/DOCKET NUMBER: 2121-102PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 205-8000
;; TELEFAX: (703) 205-8050
;; TELEX: 248345
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6555 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: circular
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: plasmid pVE144 (replicable in E.coli)
;; FEATURE:
;; NAME/KEY: -
;; LOCATION: 1..396
;; OTHER INFORMATION: /label= pUC18
;; OTHER INFORMATION: /note= "pUC18 derived sequence"
;; FEATURE:
;; NAME/KEY: -
;; LOCATION: complement (397..751)
;; OTHER INFORMATION: /label= 3'nos
;; OTHER INFORMATION: /note= "3' regulatory sequence containing the
;; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
;; OTHER INFORMATION: T-DNA nopaline synthase gene"
;; FEATURE:
;; NAME/KEY: -
;; LOCATION: complement (752..1024)
;; OTHER INFORMATION: /label= barstar
;; OTHER INFORMATION: /note= "coding region of the barstar gene of
;; OTHER INFORMATION: Bacillus amyloliquefaciens"
;; FEATURE:
;; NAME/KEY: -
;; LOCATION: complement (1025..1607)
;; OTHER INFORMATION: /label= TA29
;; OTHER INFORMATION: /note= "promoter derived from the TA29 gene of
;; OTHER INFORMATION: Nicotiana tabacum"
;; FEATURE:
;; NAME/KEY: -
;; LOCATION: 1608..2440
;; OTHER INFORMATION: /label= 35S3
;; OTHER INFORMATION: /note= "35S3 promoter sequence derived from
;; OTHER INFORMATION: cauliflower mosaic virus isolate Cab0B-J1"
;; FEATURE:
```

```
;; NAME/KEY: -
;; LOCATION: 2441..3256
;; OTHER INFORMATION: /label= neo
;; OTHER INFORMATION: /note= "coding region of the neomycine
;; OTHER INFORMATION: phosphotransferase gene of Tn5"
;; FEATURE:
;; NAME/KEY: -
;; LOCATION: 3257..4315
;; OTHER INFORMATION: /label= 3'ocs
;; OTHER INFORMATION: /note= "3' regulatory sequence containing the
;; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
;; OTHER INFORMATION: T-DNA octopine synthase gene"
;; FEATURE:
;; NAME/KEY: -
;; LOCATION: 4316..6555
;; OTHER INFORMATION: /label= pUC18
;; OTHER INFORMATION: /note= "pUC18 derived sequence"
;; US-09-025-583-2
;;
;; Query Match 100.0%; Score 24; DB 2; Length 6555;
;; Best Local Similarity 100.0%; Pred. No. 0.0057;
;; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;;
;; QY 1 tcagaagtatcagcgacctccacc 24
;; ||||||||||||||||||||||||
;; Db 993 TCAGAGTATCAGCGACCTCCACC 970
;;
;; RESULT 11
;; US-08-403-388-1
;; Sequence 1, Application US/08403388
;; Patent No. 5587289
;; GENERAL INFORMATION:
;; APPLICANT: Lurquin, Christophe; Boon-Falleur, Thierry
;; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
;; MEMBERS OF THE Mage-Xp Family and Uses Thereof
;; NUMBER OF SEQUENCES: 12
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Felfe & Lynch
;; STREET: 805 Third Avenue
;; CITY: New York City
;; STATE: New York
;; ZIP: 10022
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;; COMPUTER: IBM
;; OPERATING SYSTEM: PC-DOS
;; SOFTWARE: Wordperfect
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/403,388
;; FILING DATE: 14-MARCH-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hanson, No. 5587289man D.
;; REGISTRATION NUMBER: 30,946
;; REFERENCE/DOCKET NUMBER: LUD 5408
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 688-9200
;; TELEFAX: (212) 838-3884
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1866 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-403-388-1
;;
;; Query Match 69.2%; Score 16.6; DB 1; Length 1866;
;; Best Local Similarity 82.6%; Pred. No. 21;
;; Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Qy 2 cagaagtatcagcgacctccacc 24  
||||| |||| || |||||  
Db 498 CAGAAGCCTCAGGGAGCTCCACC 520

## RESULT 12

US-08-658-578-1  
; Sequence 1, Application US/08658578  
; Patent No. 5759783  
; GENERAL INFORMATION:  
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;  
; APPLICANT: Boon-Falleur, Thierry  
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are  
; TITLE OF INVENTION: Members Of The Mage-Xp Family and Uses Thereof  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
; COMPUTER: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/658,578  
; FILING DATE: 5-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/403,388  
; FILING DATE: 14-MARCH-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, No. 5759783man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 5444  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1866 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-658-578-1

Query Match 69.2%; Score 16.6; DB 1; Length 1866;  
Best Local Similarity 82.6%; Pred. No. 21;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24  
||||| |||| || |||||  
Db 498 CAGAAGCCTCAGGGAGCTCCACC 520

## RESULT 13

US-08-846-111D-1  
; Sequence 1, Application US/08846111D  
; Patent No. 6017705  
; GENERAL INFORMATION:  
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;  
; APPLICANT: Boon-Falleur, Thierry  
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are  
; TITLE OF INVENTION: Members Of The Mage-B Family and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York

; ZIP: 10022-7513  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
; COMPUTER: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect For DOS 6.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/846,111D  
; FILING DATE: 25-APRIL-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/658,578  
; FILING DATE: 5-MAY-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/403,388  
; FILING DATE: 14-MARCH-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, No. 6017705man D.  
; REGISTRATION NUMBER: 30,946  
; REFERENCE/DOCKET NUMBER: LUD 5444.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1866 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-846-111D-1

Query Match 69.2%; Score 16.6; DB 3; Length 1866;  
Best Local Similarity 82.6%; Pred. No. 21;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24  
||||| |||| || |||||  
Db 498 CAGAAGCCTCAGGGAGCTCCACC 520

## RESULT 14

US-09-056-105-19  
; Sequence 19, Application US/09056105  
; Patent No. 6287569  
; GENERAL INFORMATION:  
; APPLICANT: KIPPS, THOMAS J.  
; APPLICANT: WU, YUNQI  
; TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR  
; TITLE OF INVENTION: PROCESSING  
; FILE REFERENCE: 233/221  
; CURRENT APPLICATION NUMBER: US/09/056,105  
; CURRENT FILING DATE: 1998-04-06  
; EARLIER APPLICATION NUMBER: 60/043,467  
; EARLIER FILING DATE: 1997-04-10  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 19  
; LENGTH: 1866  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-056-105-19

Query Match 69.2%; Score 16.6; DB 4; Length 1866;  
Best Local Similarity 82.6%; Pred. No. 21;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24  
||||| |||| || |||||  
Db 498 cagaagcctcagggagctccacc 520

```

RESULT 15
US-08-846-111D-15
; Sequence 15, Application US/08846111D
; Patent No. 6017705
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members of The Mage-B Family and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect For DOS 6.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,111D
; FILING DATE: 25-APRIL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/658,578
; FILING DATE: 5-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,388
; FILING DATE: 14-MARCH-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6017705man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5444.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40352 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-846-111D-15

```

```

Query Match          69.2%; Score 16.6; DB 3; Length 40352;
Best Local Similarity 82.6%; Pred. NO. 38;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 cagaagtatcagcgacctccacc 24
||||| ||||| ||||| |||||
Db 38322 CAGAAGCCTCAGGAGCTCCACC 38344

```

```

Search completed: February 15, 2002, 19:08:12
Job time: 14666 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2002, 18:04:03 ; Search time 9904.61 Seconds  
(without alignments)  
26.038 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24  
Sequence: 1 tcagaagtatcagcagctccacc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*  
1: em\_estfun.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estom.\*  
5: em\_estpl.\*  
6: em\_estba.\*  
7: em\_estro.\*  
8: em\_estov.\*  
9: em\_hic.\*  
10: gb\_est1.\*  
11: gb\_est2.\*  
12: gb\_hic.\*  
13: gb\_gss.\*  
14: em\_gss\_fun.\*  
15: em\_gss\_hum.\*  
16: em\_gss\_inv.\*  
17: em\_gss\_pln.\*  
18: em\_gss\_pro.\*  
19: em\_gss\_rod.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.2	80.0	639	11	BG907143
2	19.2	80.0	641	11	BG415465
3	19.2	80.0	678	13	AZ238624
4	19.2	80.0	772	10	AW448381
5	18.8	78.3	1218	11	BG674898
6	18.2	75.8	295	11	BF487936
7	18.2	75.8	536	10	BE205329
8	18.2	75.8	592	10	AW686780
9	18.2	75.8	646	10	AW774410
10	18.2	75.8	702	10	AI812306
11	17.8	74.2	364	11	BG607914
12	17.8	74.2	746	13	AQ273843

c 13	17.6	73.3	255	10	AL039972
c 14	17.6	73.3	293	11	W17577
c 15	17.6	73.3	318	10	AA340154
c 16	17.6	73.3	325	11	BG695735
c 17	17.6	73.3	343	13	CNS00N8R
c 18	17.6	73.3	389	10	AA263139
c 19	17.6	73.3	406	11	W42373
c 20	17.6	73.3	408	13	AQ218664
c 21	17.6	73.3	414	11	BF764048
c 22	17.6	73.3	445	13	AQ06167
c 23	17.6	73.3	479	10	AI155443
c 24	17.6	73.3	520	10	AA637829
c 25	17.6	73.3	540	10	AL044230
c 26	17.6	73.3	553	11	BF993930
c 27	17.6	73.3	560	11	BF851197
c 28	17.6	73.3	564	10	AA441106
c 29	17.6	73.3	574	11	BI032905
c 30	17.6	73.3	578	10	AL046687
c 31	17.6	73.3	635	10	AI722842
c 32	17.6	73.3	638	13	AZ639135
c 33	17.6	73.3	647	11	BE897763
c 34	17.6	73.3	651	11	BI032911
c 35	17.6	73.3	680	10	AU140271
c 36	17.6	73.3	718	11	C97541
c 37	17.6	73.3	733	10	BE619157
c 38	17.6	73.3	752	11	BE904863
c 39	17.6	73.3	790	13	AQ243310
c 40	17.6	73.3	832	11	BG759915
c 41	17.6	73.3	917	11	BF204607
c 42	17.6	73.3	918	11	BG752569
c 43	17.6	73.3	937	11	BF984945
c 44	17.6	73.3	948	13	CNS02M70
c 45	17.6	73.3	1021	13	CNS02LJP

#### ALIGNMENTS

#### RESULT 1

LOCUS BG907143 639 bp mRNA EST 05-JUN-2001  
DEFINITION TaLr1157E03F TaLr1 Triticum aestivum cDNA clone TaLr1157E03 3', mRNA sequence.  
ACCESSION BG907143  
VERSION BG907143.1 GI:14314819  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 639)  
AUTHORS Cloutier,S., Dong,G. and Walsh,A.  
TITLE Wheat functional genomics- Thatcher Lr1 cDNA library  
JOURNAL Unpublished (2001)  
COMMENT Contact: Dr. Sylvie Cloutier  
Cereal Research Centre, Agriculture and Agri-food Canada  
195 Dafoe Rd. Winnipeg, MB, Canada R3T 2M9  
Tel: (204) 983-2340  
Fax: (204) 983-4604  
Email: scloutier@em.agr.ca

was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end).  
Average inset size is >2.2 kb  
Plate: 157 row: E column: 03  
Seq primer: M13 Forward.  
Location/Qualifiers  
1. 639

FEATURES  
source  
/organism="Triticum aestivum"  
/cultivar="Thatcher Lr1"  
/db\_xref="taxon:4565"  
/clone="TaLr1157E03"  
/clone\_lib="TaLr1"

/tissue\_type="Leaf tissue"  
/dev\_stage="14 Days old"  
/lab\_host="E. coli XL0LR"  
/note="Vector: Lambda ZapII; mass excised in plasmid  
vector pBK-CMV (Stratagene).; Site\_1: EcoRI; Site\_2: XhoI;  
mRNA obtained from wheat NIL Thatcher Lr1 24 hours after  
inoculation with leaf rust pathogen Puccinia triticina  
race BBB carrying the avirulence gene Avr1."  
BASE COUNT 185 a 138 c 139 g 177 t  
ORIGIN

Query Match 80.0%; Score 19.2; DB 11; Length 639;  
Best Local Similarity 87.5%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 410 TCAGAAATATCAGCGACCCACC 433

RESULT 2  
BG415465/c  
LOCUS  
DEFINITION BG415465 641 bp mRNA EST 13-MAR-2001  
HVSMEK0006J13f Hordeum vulgare testa/pericarp EST library  
HVCNDA0013 (normal) Hordeum vulgare cDNA clone HVSMEK0006J13f, mRNA  
sequence.  
ACCESSION BG415465  
VERSION BG415465.1 GI:13321016  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Hordeum vulgare

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
; Triticeae; Hordeum.  
1 (bases 1 to 641)

Wing, R., Close, T. J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu  
X., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo  
T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D. W., Main, D. and  
Wood, T.  
Development of a genetically and physically anchored EST resource  
for barley genomics  
Unpublished (2000)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: AATTACCCCTCAATAAGGG  
High quality sequence stop: 618.

Location/Qualifiers  
1..641  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSMEK0006J13f"  
/clone\_lib="Hordeum vulgare testa/pericarp EST library  
HVCNDA0013 (normal)"  
/tissue\_type="testa/pericarp"  
/lab\_host="JUC121"  
/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI; For  
more details on library preparation and sequence analysis  
see http://www.genome.clemson.edu/projects/barley/ To  
order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT 176 a 130 c 134 g 201 t

ORIGIN

Query Match 80.0%; Score 19.2; DB 11; Length 641;  
Best Local Similarity 87.5%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 198 TCAGAAATATCAGCGACCCACC 221

RESULT 4  
AW448381/c  
LOCUS

Query Match 80.0%; Score 19.2; DB 13; Length 678;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 198 TCAGAAATATCAGCGACCCACC 221

RESULT 4  
AW448381/c  
LOCUS

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 110 TCAGAAATATCAGCGACCCACC 87

RESULT 3  
AZ238624  
LOCUS  
DEFINITION AZ238624 678 bp DNA GSS 15-JUN-2000  
RPCI-23-31L2-TV RPCI-23 Mus musculus genomic clone RPCI-23-31L2,  
DNA sequence.  
ACCESSION AZ238624  
VERSION AZ238624.1 GI:8551730  
KEYWORDS GSS.  
SOURCE  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 678)  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret  
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C. M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other\_GSSs: RPCI-23-31L2.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
or from Resea ch Genetics ([http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html))  
http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 31 row: L column: 2  
Seq primer: 17  
Class: BAC ends.

Location/Qualifiers  
1..678  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-31L2"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 234 a 155 c 112 g 177 t

ORIGIN

Query Match 80.0%; Score 19.2; DB 13; Length 678;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 198 TCAGAAATATCAGCGACCCACC 221

RESULT 4  
AW448381/c  
LOCUS

Query Match 80.0%; Score 19.2; DB 13; Length 678;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 198 TCAGAAATATCAGCGACCCACC 221

RESULT 4  
AW448381/c  
LOCUS

Query Match 80.0%; Score 19.2; DB 13; Length 678;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 198 TCAGAAATATCAGCGACCCACC 221

RESULT 4  
AW448381/c  
LOCUS

Query Match 80.0%; Score 19.2; DB 13; Length 678;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 198 TCAGAAATATCAGCGACCCACC 221

RESULT 4  
AW448381/c  
LOCUS

Query Match 80.0%; Score 19.2; DB 13; Length 678;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 198 TCAGAAATATCAGCGACCCACC 221

RESULT 4  
AW448381/c  
LOCUS

Query Match 80.0%; Score 19.2; DB 13; Length 678;  
Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;





Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24  
 ||| ||||| ||||| |||||

Db 218 CAGGATTATCAGCCACCTCCACC 240

RESULT 7

BE205329 536 bp mRNA EST 05-SEP-2000  
 LOCUS EST398005 KV0 Medicago truncatula cDNA clone pKV0-21E18, mRNA sequence.

ACCESSION BE205329  
 VERSION BE205329.1 GI:8748625  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 536)  
 AUTHORS VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
 TITLE ESTs from uninoculated seedling roots of Medicago truncatula  
 JOURNAL Unpublished (1999)  
 COMMENT Department of Biology  
 Texas A&M University  
 College Station, TX 77843-3258, USA  
 Tel: 409 845 7707  
 Fax: 409 845 2891  
 Email: kate@mail.bio.tamu.edu  
 Texas A&M University name: T265101e  
 TIGR sequence name: MTGBL33TK  
 More information is available at:  
 http://chrysis.tamu.edu/medicago  
 Seq primer: SKm0d (CTA gaa CTA gta gat CC).

FEATURES  
 source  
 1..536  
 /organism="Medicago truncatula"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone\_lib="KV0"  
 /clone\_lib="KV0-21E18"  
 /tissue\_type="Seedling roots"  
 /dev\_stage="Immediately prior to inoculation with Sinorhizobium meliloti (0 hour)"  
 /lab\_host="E.coli strain XL0LR"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 140 a 175 c 101 g 120 t

ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 536;  
 Best Local Similarity 87.0%; Pred. No. 3.9e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24  
 ||| ||||| ||||| |||||

Db 314 CAGGATTATCAGCCACCTCCACC 336

RESULT 8

AW686780 592 bp mRNA EST 20-DEC-2000  
 LOCUS EST335561 KV3 Medicago truncatula cDNA clone pKV3-22024, mRNA sequence.

ACCESSION AW774410  
 VERSION AW774410.1 GI:7718325  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 592)  
 AUTHORS VandenBosch,K., Hur,J., Moore,J., Beremand,P., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
 TITLE ESTs from roots of Medicago truncatula after Rhizobium inoculation  
 JOURNAL Unpublished (1999)

DEFINITION NF042G10NR1F1000 Nodulated root Medicago truncatula cDNA clone

ACCESSION NF042G10NR 5', mRNA sequence.

VERSION AW686780 2 GI:11931346

KEYWORDS EST.

SOURCE barrel medic.  
 ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 592)  
 AUTHORS Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Iman,J.F., Weller,J.W., May G.D. and Paiva,N.L.  
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula nodulated root library  
 JOURNAL Unpublished (2000)  
 COMMENT On Apr 14, 2000 this sequence version replaced gi:7561516.  
 Contact: Paiva NL  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7317  
 Fax: 580 221 7380  
 Email: nlpaiva@noble.org  
 Insert Length: 676 Std Error: 0.00  
 Plate: 042 row: G column: 10  
 Seq primer: TCACACGAGAAACAGCTATGAC.

FEATURES  
 source  
 Location/Qualifiers  
 1..592  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone\_lib="NF042G10NR"  
 /clone\_lib="Nodulated root"  
 /tissue\_type="root"  
 /dev\_stage="Pooled developmental"  
 /note="Vector: Lambda Zap; Four-week old Rhizobium meliloti-inoculated Medicago truncatula roots, containing a mixture of young and old roots and nodules."

BASE COUNT 156 a 150 c 139 g 147 t

ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 592;  
 Best Local Similarity 87.0%; Pred. No. 4e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24  
 ||| ||||| ||||| |||||

Db 107 CAGGATTATCAGCCACCTCCACC 129

RESULT 9

AW774410 646 bp mRNA EST 07-SEP-2000  
 LOCUS EST335561 KV3 Medicago truncatula cDNA clone pKV3-22024, mRNA sequence.

ACCESSION AW774410  
 VERSION AW774410.1 GI:7718325  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 646)  
 AUTHORS VandenBosch,K., Hur,J., Moore,J., Beremand,P., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
 TITLE ESTs from roots of Medicago truncatula after Rhizobium inoculation  
 JOURNAL Unpublished (1999)

DEFINITION NF042G10NR1F1000 Nodulated root Medicago truncatula cDNA clone

ACCESSION NF042G10NR 5', mRNA sequence.

VERSION AW686780 2 GI:11931346

KEYWORDS EST.

SOURCE barrel medic.  
 ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 592)  
 AUTHORS Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Iman,J.F., Weller,J.W., May G.D. and Paiva,N.L.  
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula nodulated root library  
 JOURNAL Unpublished (2000)  
 COMMENT On Apr 14, 2000 this sequence version replaced gi:7561516.  
 Contact: Paiva NL  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7317  
 Fax: 580 221 7380  
 Email: nlpaiva@noble.org  
 Insert Length: 676 Std Error: 0.00  
 Plate: 042 row: G column: 10  
 Seq primer: TCACACGAGAAACAGCTATGAC.

FEATURES  
 source  
 Location/Qualifiers  
 1..592  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone\_lib="NF042G10NR"  
 /clone\_lib="Nodulated root"  
 /tissue\_type="root"  
 /dev\_stage="Pooled developmental"  
 /note="Vector: Lambda Zap; Four-week old Rhizobium meliloti-inoculated Medicago truncatula roots, containing a mixture of young and old roots and nodules."

BASE COUNT 156 a 150 c 139 g 147 t

ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 592;  
 Best Local Similarity 87.0%; Pred. No. 4e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24  
 ||| ||||| ||||| |||||

Db 107 CAGGATTATCAGCCACCTCCACC 129

RESULT 9

AW774410 646 bp mRNA EST 07-SEP-2000  
 LOCUS EST335561 KV3 Medicago truncatula cDNA clone pKV3-22024, mRNA sequence.

ACCESSION AW774410  
 VERSION AW774410.1 GI:7718325  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 646)  
 AUTHORS VandenBosch,K., Hur,J., Moore,J., Beremand,P., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
 TITLE ESTs from roots of Medicago truncatula after Rhizobium inoculation  
 JOURNAL Unpublished (1999)

## COMMENT

Contact: VandenBosch K  
Department of Biology  
Texas A&M University  
College Station, TX 77843-3258, USA  
Tel: 409 845 7707  
Fax: 409 845 2891  
Email: kate@mail.bio.tamu.edu  
Texas A&M EST name: T257829e  
TIGR sequence name: MTEAZ96TK  
More information is available at: .  
http://chrystie.tamu.edu/medicago  
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).

## FEATURES

source

1. .646

/organism="Medicago truncatula"  
/cultivar="genotype Al7"  
/db\_xref="taxon:3880"  
/clone="pKV3-22024"  
/clone\_lib="KV3"  
/tissue\_type="Seedling roots"  
/dev\_stage="3 days post-inoculation with Sinorhizobium meliloti"  
/lab\_host="E. coli strain XL0LR"  
/note="vector: pBluescript SK -; Site\_1: EcoRI; Site\_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT

ORIGIN

160 a 205 c 134 g 147 t

Query Match

75.8%; Score 18.2; DB 10; Length 646;

Best Local Similarity

87.0%; Pred. No. 4.1e+02;

Matches

20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

2 cagaagtacgcagcaccctccacc 24

||||| ||||| ||||| |||||

Db

283 CAGAAATTATCAGCACACCACC 305

RESULT 10

A1812306

LOCUS

A1812306 702 bp mRNA

EST

08-JUL-1999

DEFINITION

10E5 Pine Lambda Zap Xylem library Pinus taeda cDNA, mRNA sequence.

ACCESSION

A1812306

VERSION

A1812306.1 GI:5423630

KEYWORDS

EST.

SOURCE

loblolly pine.

ORGANISM

Pinus taeda

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus. 1 (bases 1 to 702)  
Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.  
The Pine Gene Discovery Project  
Unpublished (1999)  
Contact: Ross Whetten  
Forest Biotechnology Group  
North Carolina State University  
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh NC, 27695-8008  
Tel: 919-515-7800  
Fax: 919-515-7801  
Email: rosswhet@unity.ncsu.edu  
Seq primer: T3.

FEATURES

source

1. .702

/organism="Pinus taeda"  
/db\_xref="taxon:3352"  
/clone\_lib="Pine Lambda zap Xylem library"  
/tissue\_type="differentiating xylem"

/note="Vector: Lambda Zap; Site\_1: EcoRI; Site\_2: XhoI; Differentiating xylem was collected from the main stem of a 35-year old loblolly pine tree harvested during the growing season. RNA isolation and library preparation followed the methods of Allona et al., PNAS 95:9693-8, 1998"

BASE COUNT

ORIGIN

214 a 160 c 169 g 158 t 1 others  
Query Match 75.8%; Score 18.2; DB 10; Length 702;  
Best Local Similarity 87.0%; Pred. No. 4.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

2 cagaagtacgcagcaccctccacc 24

||||| ||||| ||||| |||||

Db

336 CAGAAATTACGACGCCCTCCACC 358

RESULT 11

BG607914/c

LOCUS

BG607914

DEFINITION

WHE2475\_C06\_E112S

Triticum monococcum

cDNA clone

WHE2475\_C06\_E11, mRNA

sequence.

ACCESSION

BG607914

VERSION

BG607914.1

KEYWORDS

EST.

SOURCE

Triticum monococcum.

ORGANISM

Triticum monococcum

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2001)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

Location/Qualifiers

1. .364

/organism="Triticum monococcum"

/cultivar="DV92"

/db\_xref="taxon:4568"

/clone="WHE2475\_C06\_E11"

/clone\_lib="Triticum monococcum early reproductive apex

cDNA library"

/tissue\_type="Early reproductive apex"

/dev\_stage="Seven week-old plants"

/lab\_host="E. coli XL0LR"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;

Site\_1: EcoRI; Site\_2: XhoI; The tissue, total RNA, and

poly(A) RNA were prepared from apex at double-ridge stage

to terminal-spikelet stage during transition from

vegetative state to flower state, a cDNA library was made,

and the cDNA clones were in vivo excised at the

University of California, Davis (V. Echenique, B. Stamova

, J. Dubcovsky). Plasmid DNA preparations and DNA

sequencing were performed in the Olin Anderson lab (all

other authors)."

63 a

83 c

144 g

74 t

BASE COUNT



REFERENCE  
AUTHORS

1 (bases 1 to 293)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

## TITLE

The WashU-HMI Mouse EST Project

JOURNAL  
COMMENT

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:216594

Putative full length read

Seq primer: mob.REGA+ET

High quality sequence stop: 283.

## FEATURES

Location/Qualifiers

source

1..293  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:335194"  
 /dev\_stage="19.5 dpc total fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="vector: p7T3D (Pharmacia) with a modified  
 polylinker; Site\_1: Not I ; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5'  
 TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified p7T3 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo. RNA was kindly provided by  
 Dr. Minoru Ko (Wayne State University)."

BASE COUNT  
ORIGIN

66 a 64 c 89 g 74 t

## Query Match

Best Local Similarity 73.3%; Score 17.6; DB 11; Length 293;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24

||||| | | | | | | | | |

Db 255 TCAGAAGCAGCAGTGACCCCCACC 232

## RESULT 15

## AA340154/c

## LOCUS

AA340154 318 bp mRNA EST 21-APR-1997

EST45318 Fetal brain II Homo sapiens cDNA 5' end similar to

dynactin, mRNA sequence.

AA340154

AA340154.1 GI:1992393

EST.

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 318)

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult  
 ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White  
 ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,  
 Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald  
 ,L.M., Fitzhugh,W.M., Frichman,J.L., Geoghagen,N.S., Glodok,A.,  
 Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S., Jr., Kelley,J.M.,  
 Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,

## TITLE

Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

Other ESTs: THC185498

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

## FEATURES

Location/Qualifiers

source

1..318  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):141789"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Fetal brain II"  
 /sex="female"  
 /dev\_stage="fetus, 24 wks"  
 /note="Organ: brain; Vector: pBluescript SK-; Site\_1:  
 EcoRI; Site\_2: EcoRI"

## BASE COUNT

84 a 63 c 108 g 62 t 1 others

## Query Match

Best Local Similarity 73.3%; Score 17.6; DB 10; Length 318;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24

||||| | | | | | | | | |

Db 42 TCAGCAGTATCAGCCATCTCTCTCC 19

Search completed: February 15, 2002, 18:04:07

Job time: 20956 sec

